

# SPDE Approximation for Random Trees

Yuri Bakhtin\*

## Abstract

We consider the genealogy tree for a critical branching process conditioned on non-extinction. We enumerate vertices in each generation of the tree so that for each two generations one can define a monotone map describing the ancestor–descendant relation between their vertices. We show that under appropriate rescaling this family of monotone maps converges in distribution in a special topology to a limiting flow of discontinuous monotone maps which can be seen as a continuum tree. This flow is a solution of an SPDE with respect to a Brownian sheet.

Keywords: random trees, stochastic flows, SPDE

## 1 Introduction

In this paper we present a point of view at large random trees. In [Bak] we considered Boltzmann–Gibbs distributions on rooted plane trees with bounded branching and proved that as the order of the tree grows to infinity, these trees obey a certain thermodynamic limit theorem with a limit given by an infinite Markov random tree that can be seen as the genealogy tree for a specially chosen critical branching processes conditioned on non-extinction, see [AP98],[Kes86],[Ken75]. Appropriately rescaled generation sizes in the limiting tree admit an approximation by a diffusion process. However, this view ignores all the interesting details concerning the complicated way different generations of the random tree are connected to each other.

The goal of this paper is to extend the diffusion approximation result and investigate the fine structure of the infinite random tree. The idea is to encode the tree via a stochastic flow of monotone maps. We show that under the same rescaling, the flow of monotone maps associated to the tree

---

\*School of Mathematics, Georgia Tech, Atlanta GA, 30332-0160; email:bakhtin@math.gatech.edu, 404-894-9235 (office phone), 404-894-4409(fax)

converges in distribution to a monotone flow that can be viewed as a solution of an evolutionary SPDE with respect to a Brownian sheet. The possibility of such a result was hinted at in [Bak], and in this text we provide a rigorous treatment of the SPDE, some interesting properties of its solutions, and the convergence theorem in an appropriate metric space. The solution of SPDE that we construct can be called a continuum random tree, and it appears to be a new type of continuum random trees, unknown in the literature, although some other stochastic flows similar to ours have appeared in the literature (see, e.g., [LG99]) and there is a natural connection of our results to superprocesses that describe the evolution of mass generated by continual branching particle systems conditioned on nonextinction, see [EP90],[Eva93]. We describe this connection and the advantages of our approach in Section 9. In short, the results of this paper along with those of [Bak] show what a typical large tree looks like if drawn on the plane.

**Acknowledgment.** The author wishes to thank Tom Kurtz, Carl Mueller and Andrey Pilipenko for their comments on connection of this work to the existing literature on superprocesses. He is grateful to NSF for partial support through CAREER grant DMS-0742424.

## 2 Gibbs distributions on plane trees and thermodynamic limit

In this section we describe the background in detail. Recall that plane (or, ordered) trees are rooted trees such that subtrees at any vertex are linearly ordered.

We fix  $D \in \mathbb{N}$  and introduce  $\mathbb{T}_N = \mathbb{T}_N(D)$ , the set of all plane trees on  $N$  vertices such that the branching number (i.e. the number of children) of each vertex does not exceed  $D$ .

We assume that every admissible branching number  $i \in \{0, \dots, D\}$  is assigned an energy  $E_i \in \mathbb{R}$ , and the energy of the tree  $T \in \mathbb{T}_N$  is defined via

$$E(T) = \sum_{v \in V(T)} E_{\deg(v)} = \sum_{i=0}^D \chi_i(T) E_i,$$

where  $V(T)$  denotes the set of vertices of the tree  $T$ ,  $\deg(v)$  denotes the branching number of vertex  $v$ , and  $\chi_i(T)$  is the number of vertices of degree  $i$  in  $T$ . This energy function defines a local interaction between the vertices of the tree since the energy contribution from a single vertex depends only on its nearest neighborhood in the tree.

We fix an inverse temperature parameter  $\beta > 0$  and define a probability measure  $\mu_N$  on  $\mathbb{T}_N$  by

$$\mu_N\{T\} = \frac{e^{-\beta E(T)}}{Z_N},$$

where the normalizing factor (partition function) is

$$Z_N = \sum_{T \in \mathbb{T}_N} e^{-\beta H(T)}.$$

In [Bak], the limiting behavior of measures  $\mu_N$  as  $N \rightarrow \infty$  was studied. To recall the results of [Bak], we need more notation and terminology.

For each vertex  $v$  of a tree  $T \in \mathbb{T}_N$  its height  $h(v)$  is defined as the distance to the root of  $T$ , i.e. the length of the shortest path connecting  $v$  to the root along the edges of  $T$ . The height of a finite tree is the maximum height of its vertices.

Each rooted plane tree can be uniquely encoded as a sequence of generations. By a generation we mean a monotone (nondecreasing) map  $G : \{1, \dots, k\} \rightarrow \mathbb{N}$ , or, equivalently, the set of pairs  $\{(i, G(i)) : i = 1, \dots, k\}$  such that if  $i_1 \leq i_2$  then  $G(i_1) \leq G(i_2)$ . We denote  $|G| = k$  the number of vertices in the generation, and for any  $i = 1, \dots, k$ ,  $G(i)$  denotes  $i$ 's parent number in the previous generation.

For two generations  $G$  and  $G'$  we write  $G \triangleleft G'$  and say that  $G'$  is a continuation of  $G$  if  $G'(|G'|) \leq |G|$ . Each tree of height  $n$  can be viewed as a sequence of generations

$$1 \triangleleft G_1 \triangleleft G_2 \triangleleft \dots \triangleleft G_n \triangleleft 0,$$

where  $1 \triangleleft G_1$  means  $G_1(|G_1|) = 1$  (the 0-th generation consists of a unique vertex, the root) and  $G_n \triangleleft 0$  means that the generation  $n + 1$  is empty. Infinite sequences  $1 \triangleleft G_1 \triangleleft G_2 \triangleleft \dots$  naturally encode infinite trees.

For any plane tree  $T$  and any  $n \in \mathbb{N}$ ,  $\pi_n T$  denotes the neighborhood of the root of radius  $n$ , i.e. the subtree of  $T$  spanned by all vertices with height not exceeding  $n$ .

For any  $n$  and sufficiently large  $N$ , the map  $\pi_n$  pushes the measure  $\mu_N$  on  $\mathbb{T}_N$  forward to the measure  $\mu_N \pi_n^{-1}$  on  $S_n$ , the set of all trees with height  $n$ .

We introduce  $\rho$  and  $C$  as a unique solution of

$$\frac{1}{C} = \sum_{i=0}^D e^{-\beta E_i} \rho^i = \sum_{i=0}^D i e^{-\beta E_i} \rho^i,$$

and define

$$p_i = C e^{-\beta E_i} \rho^i, \quad i = 0, 1, \dots, D.$$

Then the vector  $(p_i)_{i=0}^D$  defines a probability distribution on  $\{0, 1, \dots, D\}$  with mean 1. This vector plays the role of a minimizer for the free energy (or large deviation rate function) associated to this model, see [BH08] and [BH09].

Some of the results of [Bak] are summarized in the next theorem.

**Theorem 1** *1. There is a unique measure  $P$  on infinite rooted plane trees with branching bounded by  $D$  such that for any  $n \in \mathbb{N}$*

$$\mu_N \pi_n^{-1} \xrightarrow{TV} P \pi_n^{-1}.$$

*2. Measure  $P$  defines a Markov chain on generations  $(G_n)_{n \geq 0}$ . The transition probabilities are given by*

$$P\{G_{n+1} = g' \mid G_n = g\} = \begin{cases} \frac{|g'|}{|g|} p_{i_1} \cdots p_{i_{|g|}}, & g \triangleleft g', \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

*where  $i_k, k = 1, \dots, |g|$  denotes the number of vertices in generation  $g'$  that are children of  $k$ -th vertex in generation  $g$ .*

**Remark 1** *Our notation differs from the notation used in [Bak]. It is an easy exercise to check that the r.h.s. of (1) equals the expression given in [Bak].*

**Remark 2** *The infinite Markov tree from Theorem 1 can be obtained as a genealogy tree of a critical branching process conditioned on nonextinction, see [AP98], [Kes86], and [Ken75].*

Let us also recall a functional limit theorem for  $X_n = |G_n|$ , the process of the Markov infinite tree's generation sizes. We introduce moments of the distribution  $p$ :

$$B_n = \sum_{i=0}^D i^n p_i, \quad n \in \mathbb{N}, \quad (2)$$

and its variance

$$\mu = B_2 - B_1^2 = B_2 - 1.$$

Let us now fix a positive time  $T$  and define

$$Z_n(t) = \frac{1}{\mu n} (X_{[nt]} + \{nt\} (X_{[nt+1]} - X_{[nt]})), \quad t \in [0, T],$$

the rescaled process of linear interpolation between values of  $X_k$  given at whole times  $k = 0, 1, \dots$

**Theorem 2** *The distribution of  $Z_n$  converges weakly in uniform topology on  $C([0, T])$  to the distribution of a diffusion process with generator*

$$Lf(x) = (f'(x) + \frac{1}{2}xf''(x))\mathbf{1}_{\{x \geq 0\}}, \quad (3)$$

*emitted from 0 at time 0.*

Notice that the limiting process can be viewed as a weak solution of an SDE

$$dX(t) = dt + \sqrt{X(t)}dW(t)$$

on  $\mathbb{R}_+$  (zero is entrance and non-exit singular boundary point for  $\mathbb{R}_+$ ).

This result captures the asymptotics of a rather rough characteristic of the tree, the generation size. It ignores all the interesting details of the way the generations are connected to each other. In this paper, we are interested in taking these details into account, thus refining the diffusion approximation.

From now on we consider the random genealogy trees defined above via the probability vector  $(p_i)_{i=0}^D$ . Although  $D$  will always be assumed to be finite, the results also hold true if one replaces the boundedness of  $D$  with certain moment restrictions on the distribution.

### 3 Limit theorems for finite partitions

Let us take a positive number  $z$ , and for each  $n \in \mathbb{N}$  consider the Markov process on generations (described in the previous section) originating at time 0 with a population of  $[nz]$  vertices. Let us take  $m \in \mathbb{N}$  and choose some partition  $0 = x_0 < x_1 < \dots < x_{m-1} < x_m = z$  of  $[0, z]$ . Let us denote by  $U(n, k, j)$  the size of the progeny of first  $[nx_k]$  vertices in the initial population after  $j$  steps. Notice that  $V(n, k, j) = U(n, k, j) - U(n, k-1, j)$  describes the evolution of the progeny of disjoint subpopulations of the original population. Denoting the rescaled interpolation by

$$V_{n,k}(t) = \frac{1}{\mu n} (V(n, k, [nt]) + \{nt\}(V(n, k, [nt]+1) - V(n, k, [nt]))), \quad t \in [0, T],$$

and following the same lines as in [Bak], we obtain the following result:

**Theorem 3** *Vector-valued process  $(V_{n,k})_{k=1}^m$  converges (as  $n \rightarrow \infty$ ) in distribution in the uniform topology to a diffusion process  $(V_k)_{k=1}^m$  on  $\mathbb{R}_+^m$  with*

initial data  $V_k(0) = v_k = x_k - x_{k-1}$  and generator  $L$ , for all  $C^2$ -functions with compact support given by

$$Lf(y) = \sum_{k=1}^m \frac{y_k}{|y|_1} \mathbf{1}_{\{y_k \geq 0\}} \partial_k f(y) + \frac{1}{2} \sum_{k=1}^m y_k \mathbf{1}_{\{y_k \geq 0\}} \partial_{kk}^2 f(y), \quad y \in \mathbb{R}_+^m, \quad (4)$$

where  $|y|_1 = y_1 + \dots + y_m$ .

Equivalently, the limiting process can be represented as a coordinatewise nonnegative solution of a system of SDEs:

$$dV_k(t) = \frac{V_k(t)}{|V(t)|_1} dt + \sqrt{V_k(t)} dW_k(t), \quad (5)$$

$$V_k(0) = v_k, \quad k = 1, \dots, m, \quad (6)$$

where  $(W_k)_{k=1}^m$  are independent Wiener processes.

It is clear that this system has a unique strong solution until the first time at which  $V_k = 0$  for some  $k$ . It is also easy to see (essentially from the Feller classification of singular points for 1-dimensional SDEs) that for each  $k = 1, \dots, m$ , the coordinate hyperplane  $\{x_k = 0\}$  is an absorbing set. Therefore, as soon as the system reaches one of the coordinate planes, we have to solve a system of  $m - 1$  equations from that point on. Iterating this process we obtain a unique strong solution in  $\mathbb{R}_+^m$ .

In the same way, a sequence  $(U_{n,k})_{k=1}^m$  defined by

$$U_{n,k}(t) = \frac{1}{\mu n} (U(n, k, [nt]) + \{nt\} (U(n, k, [nt] + 1) - U(n, k, [nt]))), \quad t \in [0, T],$$

satisfies a functional limit theorem.

**Theorem 4** *The vector-valued process  $(U_{n,k})_{k=1}^m$  converges in distribution in uniform topology to a limiting process  $(U_n)_{k=1}^m$  that can be represented as*

$$U_k(t) = V_1(t) + \dots + V_k(t), \quad k = 1, \dots, m,$$

where  $V$  is the limiting process from Theorem 3.

Equivalently, the limiting process can be represented as a (nondecreasing in  $k$ ) solution of a system of SDEs:

$$dU_k(t) = \frac{U_k(t)}{U_m(t)} dt + \sum_{j=1}^k \sqrt{U_k(t) - U_{k-1}(t)} dW_j(t), \quad (7)$$

$$U_k(0) = x_k, \quad k = 1, \dots, m, \quad (8)$$

where  $(W_j)_{j=1}^m$  are independent Wiener processes.

Although the theorems above provide interesting information about the behavior of the progeny of finitely many subpopulations, we do not give detailed proofs of these results, since we are actually aiming at more advanced ones (nevertheless, see Section 8.3 for a proof of a similar statement).

It is important to notice that the system (7),(8) (or, equivalently, (5),(6)) not only determines the evolution of process  $U$  (respectively  $V$ ) in time, but also has a rich “spatial” structure which will be explored in later sections.

## 4 SPDE approximation: first steps

In this section we give an informal view at the possibility of constructing a limiting continuum random tree represented as a solution to a certain SPDE.

In the approach described in the last section, for any partition  $(x_k)_{k=0}^m$  we were able to find a probability space and a diffusion process defined on it that serves as a weak limit for the evolution of subpopulations in a discrete tree. Note that for any  $t \geq 0$ , the pre-limit processes  $V_n, U_n$  as well as their limits  $U, V$  constructed above give rise to a random monotone map  $x_k \mapsto U_k(t)$  defined on  $(x_k)_{k=0}^m$  and the aforementioned probability space.

A drawback of this approach is that we need a separate probability space for each partition whereas it would be natural to have just one probability space and a random flow of monotone maps  $x \mapsto U(x, t)$  (defined on that space) that serves all partitions simultaneously. That is, we would like to require that for any  $m$  and any partition  $(x_k)_{k=0}^m$  of  $[0, z]$ ,  $(U(x_k, t) - U(x_{k-1}, t))_{k=1}^m$  is a Markov process with generator  $L$  defined in (4).

To that end we utilize the spatial structure that has been mentioned in the end of Section 3 and introduce a stochastic equation with respect to a Brownian sheet  $W$  defined on Borel subsets of  $\mathbb{R}_+^2 = \{(x, t) : x, t \geq 0\}$  and viewed as an orthogonal martingale measure (see [Wal86, Chapter 2] for a definition of Brownian sheet and stochastic integration with respect to orthogonal martingale measures):

$$\begin{aligned} dU(x, t) &= \frac{U(x, t)}{U(z, t)} dt + W([0, U(x, t)] \times dt), \\ U(x, t_0) &= x, \quad x \in [0, z], \end{aligned}$$

or, in integral form,

$$U(x, t) = x + \int_0^t \frac{U(x, s)}{U(z, s)} ds + \int_0^t \int_{\mathbb{R}} \mathbf{1}_{[0, U(x, t)]}(y) W(dy \times dt), \quad (9)$$

The choice of this SPDE is a natural one, since the drift term and the quadratic covariation structure of the martingale part coincide with those of (7).

Our discussion of this SPDE will concentrate around the following questions: a natural definition of a solution, its existence and uniqueness, regularity properties, and the relation to the finite-dimensional system (7),(8).

First, let us take any  $z > 0$  and  $x = z$  in (9). The equation rewrites then as

$$U(z, t) = z + t + \int_0^t \int_{\mathbb{R}} \mathbf{1}_{[0, U(z, t)]}(y) W(dy \times dt). \quad (10)$$

One can obtain a unique strong solution of this equation using the standard Picard iteration scheme and mimicking the proof of existence and uniqueness for an ordinary SDE. The only potential difficulty one may encounter is the behavior of the solution near 0, but the solution is a Markov process with generator  $L$  given in (3) so that the solution always stays positive with probability 1.

Next, we can take any  $z > 0$  and a finite set  $I$  of admissible initial points, i.e.,

$$\{0, z\} \subset I \subset [0, z].$$

Mimicking the construction and the proof of the uniqueness of a positive solution of (5),(6) we see that (9) has a unique strong nondecreasing in  $x \in I$  solution  $(U(x, t))_{x \in I, t \geq 0}$ . Thus obtained  $U$  satisfies a finite system of stochastic equations w.r.t. the Brownian sheet with probability one. It is clear that if  $I' \supset I$  is a broader finite set of admissible initial conditions, then, with probability 1, for any  $t \geq 0$ , the map  $x \mapsto U(x, t), x \in I'$  is a monotone extension of the map  $x \mapsto U(x, t), x \in I$ .

In principle, there might be a problem with defining this map with probability 1 for an uncountable set  $x \in [0, z]$ . However, we can use the monotonicity to tackle this difficulty. Let us take a countable set  $J$  of admissible initial conditions such that  $J$  is dense in  $[0, z]$ . Then, taking a sequence of finite sets increasing to  $J$ , and iteratively extending the map  $x \mapsto U(x, t)$  on each set of the sequence as described above, we are able to define this map on  $J$  with probability 1, for all  $t \geq 0$ . Moreover, this map is a.s.-monotone for all  $t$ . Therefore, this map can be extended uniquely by monotonicity to all points of  $[0, z]$  except for at most countable set of discontinuities. Notice that for two different dense sets  $J, J'$  of admissible initial conditions, with probability 1, the maps  $U$  defined on  $J$  and  $J'$  are monotone continuations of each other. Therefore the monotone extensions of these two maps agree at all points of  $[0, z]$  except for a countable set of upward jumps.



It is noteworthy that discontinuities in the form of upward jumps, or shocks, do occur with probability 1 as we shall see in Section 7. To deal with them we are going to ignore the value of  $U(x, t)$  at the jump, and work with equivalence classes of monotone functions coinciding at every continuity point. We proceed to introduce these equivalence classes in the next section.

## 5 Monotone graphs and flows

The goal of this section is to introduce a space that will serve as a natural existence and uniqueness class of solutions of SPDE (9).

Consider all points  $z \geq 0$  and nonnegative nondecreasing functions  $f$  defined on  $(-\infty, z]$  such that  $f(x) = 0$  for all  $x < 0$ . Each of these functions has at most countably many discontinuities. We say that two such functions  $f_1 : (-\infty, z_1] \rightarrow \mathbb{R}^+$ ,  $f_2 : (-\infty, z_2] \rightarrow \mathbb{R}^+$  are equivalent if  $z_1 = z_2$ ,  $f_1(z_1) = f_2(z_2)$ , and for each continuity point  $x$  of  $f_1$ ,  $f_1(x) = f_2(x)$ . Although the roles of  $f_1$  and  $f_2$  seem to be different in this definition, it is easily seen to define a true equivalence relation. The set of all classes of equivalence will be denoted by  $\mathbb{M}$ . We would like to endow  $\mathbb{M}$  with a metric structure, and to that end we develop a couple of points of view.

Sometimes, it is convenient to identify each element of  $\mathbb{M}$  with its unique right-continuous representative. Sometimes, it is also convenient to work with graphs. The graph of a monotone function  $f$  defined on  $(-\infty, z]$  is the set  $G_f = \{(x, f(x)), x \leq z\}$ . For each discontinuity point  $x$  of  $f$  one may consider the line segment  $\bar{f}(x)$  connecting points  $[x, f(x-)]$  and  $[x, f(x+)]$ . The continuous version of  $G_f$  is the union of  $G_f$  and all segments  $\bar{f}(x)$ . It is often convenient to identify an element of  $\mathbb{M}$  with a continuous version of its graph restricted to  $\mathbb{R}_+^2$ , and we shall do so from now on calling the elements of  $\mathbb{M}$  monotone graphs. Yet another way to look at monotone graphs is to think of them as monotone multivalued maps so that the image of each point is either a point  $f(x)$  or a segment  $\bar{f}(x)$ .

The Hausdorff distance between  $\Gamma_1 \in \mathbb{M}$  and  $\Gamma_2 \in \mathbb{M}$  is defined via

$$\rho(\Gamma_1, \Gamma_2) = \max \left\{ \sup_{z_1 \in \Gamma_1} \inf_{z_2 \in \Gamma_2} |z_1 - z_2|, \sup_{z_2 \in \Gamma_2} \inf_{z_1 \in \Gamma_1} |z_1 - z_2| \right\}.$$

**Lemma 1**  $(\mathbb{M}, \rho)$  is a Polish (complete and separable) metric space.

PROOF: The separability follows since one can easily approximate any monotone graph by broken lines with finitely many rational vertices.

The metric space of compact subsets of the plane is complete with respect to Hausdorff metric. Therefore, to establish the completeness of  $(\mathbb{M}, \rho)$  we

need to prove that a limit of a convergent sequence of monotone graphs is necessarily a monotone graph.

The limit set is obviously a connected one, with at most one point on each of the lines  $\{(x_0, x_1) : \alpha_0 x_0 + \alpha_1 x_1 = s\}$  for any  $\alpha_0, \alpha_1 > 0$  and any  $s \in \mathbb{R}$ . Therefore it is a curve  $\{\gamma_0(s), \gamma_1(s)\}$  parametrized by  $s = \gamma_0(s) + \gamma_1(s)$ , and the monotonicity follows easily.  $\square$

We can derive another view at the space  $\mathbb{M}$  from the proof above. We can look at each monotone graph in  $\mathbb{M}$  in a rotated coordinate frame: then we shall see that the set  $\{(\gamma_0(s) + \gamma_1(s), \gamma_1(s) - \gamma_0(s))\}$  is a graph of 1-Lipschitz function.

We can introduce another distance based on this viewpoint. Take any  $\Gamma_1, \Gamma_2 \in \mathbb{M}$  and consider these curves in  $\mathbb{R}_+^2$  in the rotated coordinate system as 1-Lipschitz functions  $g_1$  and  $g_2$  over  $[0, s_1]$  and  $[0, s_2]$  respectively. Define  $s_* = s_1 \wedge s_2$ , and

$$\rho'(\Gamma_1, \Gamma_2) = |s_1 - s_2| + \sup_{0 \leq s \leq s_*} |g_1(s) - g_2(s)|.$$

The proof of the following lemma is an easy exercise.

**Lemma 2** *Distances  $\rho$  and  $\rho'$  are equivalent on  $\mathbb{M}$ .*

The following is a useful criterion of convergence in  $(\mathbb{M}, \rho)$ .

**Lemma 3** *A sequence of monotone graphs  $(\Gamma_n)_{n \in \mathbb{N}}$  with right-continuous representatives  $f_n : [0, z_n] \rightarrow \mathbb{R}_+, n \in \mathbb{N}$ , converges in  $\rho$  to  $\Gamma$  with right-continuous representative  $f : [0, z] \rightarrow \mathbb{R}_+$  iff*

1.  $z_n \rightarrow z$ , as  $n \rightarrow \infty$ ;
2.  $f_n(z_n) \rightarrow f(z)$ , as  $n \rightarrow \infty$ ;
3. At each continuity point  $x < z$  of  $f$ ,  $f_n(x) \rightarrow f(x)$ , as  $n \rightarrow \infty$ .

PROOF: Suppose conditions 1–3 are satisfied. If  $z = 0$ , then the convergence is obvious. If  $z > 0$ , take any  $\varepsilon > 0$  and find continuity points  $x_1 < x_2 < \dots < x_m$  so that  $x_1 < \varepsilon$ ,  $x_2 - x_1 < \varepsilon, \dots, z - x_m < \varepsilon$ . Then take  $n_0 \in \mathbb{N}$  such that for all  $n > n_0$ ,

- (i) for all  $k = 1, \dots, m$ ,  $x_k < z_n$ ;
- (ii) for all  $k = 1, \dots, m$ ,  $|f_n(x_k) - f(x_k)| < \varepsilon$ ;
- (iii)  $|z - z_n| < \varepsilon$ ;

$$(iv) \quad |f_n(z) - f(z_n)| < \varepsilon.$$

It is easy to see that due to the monotonicity, for these values of  $n$  we have  $\rho(\Gamma_n, \Gamma) < 2\varepsilon$ , and thus the convergence holds.

Suppose now that the convergence holds. Conditions 1 and 2 of the lemma follow immediately by monotonicity. If condition 3 is violated, then there is  $\delta > 0$  and a subsequence  $(n')$  such that  $|f_{n'}(x) - f(x)| > \delta$ . We can choose  $\varepsilon > 0$  so that  $|y - x| < \varepsilon$  implies  $|f(x) - f(y)| < \delta/2$ . Therefore, for each  $n'$ , the point  $(x, f_{n'}(x))$  is at least at distance  $\varepsilon \wedge (\delta/2)$  from any point on  $\Gamma$  which contradicts the assumption and completes the proof.  $\square$

The criterion of convergence provided by the above lemma allows to conclude immediately that any bounded set of monotone graphs is precompact. Let us introduce  $z_j(\Gamma) = \sup\{x_j : (x_0, x_1) \in \Gamma\}$ ,  $j = 0, 1$ .

**Lemma 4** *A set  $H \in \mathbb{M}$  is precompact in  $\mathbb{M}$  iff the set  $\{(z_0(\Gamma), z_1(\Gamma)) : \Gamma \in H\}$  is bounded in  $\mathbb{R}^2$ .*

This lemma can also be derived from Lemma 2 since the Lipschitz constant for monotone graphs in the rotated coordinate system is bounded by 1.

Using Lemma 3, we can also prove the following statement, which we will not use but give here for completeness:

**Lemma 5** *For any  $z > 0$ , the set  $\mathbb{M}(z)$  of all monotone graphs  $\Gamma \in \mathbb{M}$  with  $z_0(\Gamma) = z$  is a closed set. Convergence of monotone graphs in  $\mathbb{M}(z)$  in distance  $\rho$  is equivalent to essential convergence of their monotone right-continuous representatives which is in turn equivalent to their convergence in Skorokhod topology on  $D([0, z])$ .*

For two monotone graphs  $\Gamma_1$  and  $\Gamma_2$  with  $z_1(\Gamma_1) = z_0(\Gamma_2)$ , we define their composition  $\Gamma_2 \circ \Gamma_1$  as the set of all pairs  $(x_0, x_1)$  such that  $(x_0, x_2) \in \Gamma_1$  and  $(x_2, x_1) \in \Gamma_2$  for some  $x_2$ .

Let  $T > 0$  and  $\Delta_T = \{(t_0, t_1) : 0 \leq t_0 \leq t_1 \leq T\}$ . We say that  $(\Gamma^{t_0, t_1})_{(t_0, t_1) \in \Delta_T}$  is a (continuous) *monotone flow* on  $[0, T]$  if the following properties are satisfied:

1. For each  $(t_0, t_1) \in \Delta_T$ ,  $\Gamma^{t_0, t_1}$  is a monotone graph.
2. The monotone graph  $\Gamma^{t_0, t_1}$  depends on  $(t_0, t_1)$  continuously in  $\rho$ .
3. For each  $t \in [0, T]$ ,  $\Gamma^{t, t}$  is the identity map on  $[0, Z(t)]$  for some  $Z(t)$ .  
The function  $Z$  is called the *profile* of  $\Gamma$ .

4. For any  $(t_0, t_1) \in \Delta_T$ ,  $z_0(\Gamma^{t_0, t_1}) = Z(t_0)$ ,  $z_1(\Gamma^{t_0, t_1}) = Z(t_1)$ , where  $Z$  is the profile of  $\Gamma$ .
5. If  $(t_0, t_1) \in \Delta_T$  and  $(t_1, t_2) \in \Delta_T$ , then  $\Gamma^{t_0, t_2} = \Gamma^{t_1, t_2} \circ \Gamma^{t_0, t_1}$ .

It is easy to check that the space  $\mathbb{M}[0, T]$  of all monotone flows on  $[0, T]$  is a Polish space if equipped with uniform Hausdorff distance:

$$\rho_T(\Gamma_1, \Gamma_2) = \sup_{(t_0, t_1) \in \Delta_T} \rho(\Gamma_1^{t_0, t_1}, \Gamma_2^{t_0, t_1}). \quad (11)$$

Property 5 (consistency) implies that Property 2 (continuity) has to be checked only for  $t_0 = t_1$ .

Our next goal is to introduce trajectories of individual points in the monotone flow. Suppose  $\Gamma \in \Delta_T$ , and  $Z$  is the profile of  $\Gamma$ . Let  $U : \mathbb{R}^+ \times \Delta_T \rightarrow \mathbb{R}_+$  satisfy the following properties:

1. For any  $(t_0, t_1) \in \Delta_T$ , the function  $U(x, t_0, t_1)$  is monotone in  $x \in [0, Z(t_0)]$ .
2. For any  $(t_0, t_1) \in \Delta_T$ , if  $x \in [0, Z(t_0)]$ , then  $(t_1, U(x, t_0, t_1)) \in \Gamma^{t_0, t_1}$ .
3. For all  $x, t_0$ ,  $U(x, t_0, t_1)$  is continuous in  $t_1$ .

Then  $U$  and  $\Gamma$  are said to be compatible with each other, and  $U$  is said to be a *trajectory representation* of  $\Gamma$ . Clearly, the monotonicity implies that, given  $Z$ , there is at most one monotone flow on  $[0, T]$  compatible with  $U$ . Moreover, it is sufficient to know a trajectory representation  $U(x, t_0, t_1)$  for a dense set of points  $x, t_0, t_1$  (e.g., rational points) to reconstruct the flow.

Although a trajectory representation for a monotone flow  $\Gamma$  with profile  $Z$  is not unique, there is a special representation  $U(x, t_0, t_1)$  that is right-continuous in  $x \in [0, Z(t_0)]$  for every  $t_1 \geq t_0$ :

$$U(x, t_0, t_1) = \begin{cases} \sup\{y : (x, y) \in \Gamma^{t_0, t_1}\}, & x \in [0, Z(t_0)] \\ x, & x > Z(t_0). \end{cases} \quad (12)$$

The concrete way of defining  $U(x, t_0, t_1)$  for  $x > Z(t_0)$  is inessential for our purposes, and we often will simply ignore points  $(x, t_0, t_1)$  with  $x > Z(t_0)$ .

It is often convenient to understand a monotone flow as a triple  $\Gamma = (\Gamma, Z, U)$ , where  $Z$  is the profile of  $\Gamma$ , and  $U$  is one of the trajectory representations of  $\Gamma$ .

We are now ready to define a solution of our main equation (9).

## 6 Solution of the SPDE

In this section, we define a solution of (9), prove its existence, uniqueness, and the Markov property.

Let  $(\Omega, \mathcal{F}, \mathbf{P})$  be a complete probability space with a filtration  $(\mathcal{F}_t)_{t \geq 0}$  satisfying the usual conditions (right-continuity and completeness), and let  $W$  be a Brownian sheet on  $\mathbb{R}_+^2$  w.r.t.  $(\mathcal{F}_t)_{t \geq 0}$ , i.e.,  $W$  is a centered Gaussian random field indexed by Borel subsets of  $\mathbb{R}_+^2$ , such that for any Borel set  $A \subset \mathbb{R}_+$  with finite Lebesgue measure  $|A|$ ,  $(W([0, t] \times A))_{t \geq 0}$ , is an  $(\mathcal{F}_t)_{t \geq 0}$ -martingale with  $\langle W([0, \cdot] \times A) \rangle_t = t|A|$ . We refer to [Wal86] for more background on martingale measures.

We define a solution to (9) on  $[0, T]$  as a random monotone flow  $\Gamma : \Omega \rightarrow \mathbb{M}[0, T]$  such that with probability 1, the triplet  $(\Gamma, Z, U)$  has the following properties for some trajectory representation  $U$  of  $\Gamma$ :

1. For any  $t_0 \geq 0$  and any  $t_1 \geq t_0$ ,  $\Gamma^{t_0, t_1}$  is measurable w.r.t.  $\mathcal{F}_{t_1}$ .
2. For any  $x \geq 0$  and  $(t_0, t_1) \in \Delta$ ,  $U(x, t_0, t_1)$  is measurable w.r.t.  $\mathcal{F}_{t_1}$ .
3. For any  $t_0 \geq 0$ , any  $x \geq 0$ , almost every  $\omega \in \{x \leq Z(t_0)\}$ , and all  $t_1 \geq t_0$ ,

$$U(x, t_0, t_1) = x + \int_{t_0}^{t_1} \frac{U(x, t_0, t)}{Z(t)} dt + \int_{t_0}^{t_1} \int_{\mathbb{R}} \mathbf{1}_{[0, U(x, t_0, t)]}(y) W(dy \times dt). \quad (13)$$

Let us recall that the idea of this definition is to define a natural existence and uniqueness class of solutions by means of consistent co-evolution of monotone graphs and trajectories of individual points.

**Theorem 5** *For any filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$  satisfying the usual conditions and a Brownian sheet  $W$  w.r.t.  $(\mathcal{F}_t)$ , there is a unique solution  $(\Gamma^{t_0, t_1})_{(t_0, t_1) \in \Delta}$  of equation (9) in the above sense.*

PROOF: Let us construct a solution first. We begin with  $Z$ . Let us start with an auxiliary equation

$$Z(x, t) = x + t + \int_0^t \int_{\mathbb{R}} \mathbf{1}_{[0, Z(x, s)]}(y) W(dy \times ds)$$

for  $x > 0$ . Mimicking the SDE methodology, it is easy to construct a unique strong solution of this equation defined up to a stopping time

$$\tau_n = \inf \left\{ t \geq 0 : Z(x, t) \leq \frac{1}{n} \right\}.$$

Letting  $n \rightarrow \infty$  and noticing that 0 is a no-exit singular point for the diffusion  $Z(x, \cdot)$ , we can extend this solution to the whole time semi-axis  $\mathbb{R}_+$ . Using the uniqueness, we see that with probability 1, for all  $m \in \mathbb{N}$  and  $t \geq 0$ ,

$$Z(1/m, t) \geq Z(1/(m+1), t),$$

so that the limit  $Z(t) = \lim_{m \rightarrow \infty} Z(1/m, t)$  is well-defined. It is easy to check now that  $Z(t)$  is a strong solution of equation

$$Z(t) = t + \int_0^t \int_{\mathbb{R}} \mathbf{1}_{[0, Z(s)]}(y) W(dy \times dt).$$

It is also easy to check that this strong solution is a unique one by mimicking the proof of the uniqueness theorem of Yamada and Watanabe for SDEs, see Proposition 2.13 in [KS88, Chapter 5])

For any two nonnegative rational numbers  $x, t$ , we define a stochastic process  $(U(x, t_0, t_1))_{t_1 \geq t_0}$  as follows. If  $x > Z(t_0)$ , we set  $U(x, t_0, t_1) = x$ ; if  $x \leq Z(t_0)$ , we set  $U(x, t_0, \cdot)$  to be the unique strong solution of equation (13) on  $[t_0, \infty)$  (the uniqueness can be established easily using methods borrowed from the theory of SDEs).

For each  $t_1 \geq t_0$  the resulting map  $x \mapsto U(x, t_0, t_1)$  is nondecreasing on  $[0, Z(t_0)]$ : if  $x_1 \leq x_2$  then  $U(x_1, t_0, t_1) \leq U(x_2, t_0, t_1)$ . In fact, if the opposite inequality holds for some  $t_1 > t_0$  then  $U(x_1, t_0, t') = U(x_2, t_0, t')$  for some  $t' \in [t_0, t_1]$ , and the uniqueness implies that  $U(x_1, t_0, \cdot)$  and  $U(x_2, t_0, \cdot)$  coincide after  $t'$ .

Now, for any  $(t_0, t_1) \in \Delta$  with rational  $t_0$ , the (random) map  $x \rightarrow U(x, t_0, t_1)$  defined above uniquely determines a monotone graph  $\Gamma^{t_0, t_1}$ . Using the continuity of trajectories  $U(x, t_0, t_1)$  and their monotonicity property, it is easy to show that  $\Gamma^{t_0, t_1}$  depends continuously on rationals  $t_0$  and  $t_1$ . Therefore, we can extend  $\Gamma^{t_0, t_1}$  by continuity to all values  $(t_0, t_1) \in \Delta$ .

Knowing monotone functions  $U(x, t_0, t_1)$  for rational values of  $x$  and  $t_0$ , and all  $t_1 \geq t_0$  allows to define

$$U(x, t_0, t_1) = \inf\{U(y, t, t_1) : t, y \in \mathbb{Q}, t < t_0, y \in [0, Z(t)], U(y, t, t_0) > x\},$$

for all other values of  $t_0$  and  $x \in [0, Z(t_0)]$ , and it is easy to verify that thus defined random field  $U$  satisfies the requirement of the Theorem.

To prove the uniqueness, it suffices to notice that the result of each step in the above construction of is a.s.-uniquely defined.  $\square$

The proof of the theorem provides a construction of a unique solution of SPDE (9). In the next section, we study some properties of the solution.

**Remark 3** *The set of all pairs of rationals  $x, t$  in the proof of Theorem 5 can be replaced by any countable and dense set in  $\mathbb{R}_+^2$*

The following statement follows directly from the Markov property of  $U(x, t_0, t_1)$  in  $t_1$ :

**Lemma 6 (Markov property)** *For any  $t_0 \geq 0$ ,  $(\Gamma^{t_0, t_1}, t_1 \geq t_0)$  is a Markov process.*

## 7 Discontinuities of the solution

The reason for introducing monotone flows the way we did in Sections 5 and 6 is the presence of discontinuities of solutions with respect to  $x \in [0, Z(t)]$ . If not for these discontinuities our analysis would have been easier, and in this section we will show that they are, in fact, intrinsic to the solution, thus justifying our choice to work with monotone graphs.

Let us call a straight line segment connecting two points  $(x_1, y_1)$  and  $(x_2, y_2)$  on the plain vertical if  $x_1 = x_2$ .

**Lemma 7** *Let  $0 \leq t_0 < t_1$ . Then, with probability 1,  $\Gamma^{t_0, t_1}$  contains nondegenerate vertical segments, i.e., any monotone function representing  $\Gamma^{t_0, t_1}$  is discontinuous, with shocks associated to these vertical segments.*

PROOF: To prove this lemma we need the following standard result (see e.g. the proof of Theorem 26 in [Pro04, Chapter 2] for this result in the context of realizations of stochastic processes).

**Lemma 8** *If  $f : [0, z] \rightarrow \mathbb{R}$  is a bounded variation function then its quadratic variation  $Q(f)$  defined by*

$$Q(f) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} (f(z(i+1)/n) - f(zi/n))^2$$

*satisfies:*

$$Q(f) = \sum_{x \in \Delta(f)} (f(x+) - f(x-))^2,$$

*where  $\Delta(f)$  is the set of all discontinuity points of  $f$ . In particular,  $f$  is continuous on  $[0, z]$  iff  $Q(f) = 0$ .*

Let us compute the quadratic variation of  $U(\cdot, t_0, t_1)$ . For  $n \in \mathbb{N}$  we introduce (omitting dependence on  $n$ )

$$x_k = \frac{k}{n}Z(t_0), \quad k = 0, \dots, n,$$

and

$$V_k(t_1) = U(x_k, t_0, t_1) - U(x_{k-1}, t_0, t_1), \quad k = 1, \dots, n.$$

Itô's formula implies

$$\begin{aligned} dV_k^2(t_1) &= 2V_k(t_1)dV_k(t_1) + V_k(t_1)dt_1 \\ &= \frac{2V_k^2(t_1)}{Z(t_1)}dt_1 + 2V_k(t_1)W([U(x_{k-1}, t_0, t_1), U(x_k, t_0, t_1)] \times dt_1) \\ &\quad + V_k(t_1)dt_1. \end{aligned}$$

Let

$$Q_n(t_1) = V_1^2(t_1) + \dots + V_n^2(t_1).$$

Then  $Q_n(t_0) = (Z(t_0))^2/n$ , and

$$\begin{aligned} Q_n(t_1) &= \frac{(Z(t_0))^2}{n} + 2 \int_{t_0}^{t_1} \frac{Q_n(t)}{Z(t)} dt \\ &\quad + 2 \sum_{k=1}^n \int_{t_0}^{t_1} V_k(t)W([U(x_{k-1}, t_0, t), U(x_k, t_0, t)] \times dt) \\ &\quad + \int_{t_0}^{t_1} Z(t)dt. \end{aligned}$$

Let us define  $Q(t) = \lim_{n \rightarrow \infty} Q_n(t)$  and  $\nu = \inf\{t > t_0 : Q(t) > 0\}$ . If  $\nu > t_0$ , then taking the limit as  $n \rightarrow \infty$  in both sides of the equation above at  $t_1 = \nu$ , we see that all terms converge to zero except for  $\int_{t_0}^{\nu} Z(t)dt$ . To obtain the convergence to zero for the martingale stochastic integral term, it is sufficient to see that the quadratic variation in time given by

$$\left\langle \sum_{k=1}^n \int_{t_0}^{\cdot} V_k(t)W([U(x_{k-1}, t_0, t), U(x_k, t_0, t)] \times dt) \right\rangle_{\nu} = \int_{t_0}^{\nu} \sum_{k=1}^n V_k^2(t)dt,$$

converges to zero as  $n \rightarrow \infty$ .

Since  $\int_{t_0}^{\nu} Z(t)dt$  is a strictly positive random variable that does not depend on  $n$ , we obtain a contradiction which shows that  $\nu = t_0$ , so that  $Q(t) > 0$  for any  $t > t_0$ .  $\square$



## 8 Convergence

This section is the central part of the paper. Here we prove that the discrete infinite random genealogy tree  $\tau$  converges in distribution in an appropriate sense under appropriate rescaling to the continuum random tree given by the monotone flow described in the previous section.

### 8.1 Trees as monotone flows. Main result.

To start with, we introduce a monotone flow associated with a realization of the infinite discrete tree  $\tau$ . This procedure is analogous to that of linear interpolation for discrete time random walks leading to Donsker's invariance principle.

We begin with an imbedding of the tree in the plane. Recall that there are  $X_n \geq 1$  vertices in the  $n$ -th generation of the tree. For  $i \in \{1, \dots, X_n\}$ , the  $i$ -th vertex of  $n$ -th generation is represented by the point  $(n, i - 1)$  on the plane. The parent-child relation between two vertices of the tree is represented by a straight line segment connecting the representations of these vertices.

Besides these “regular” segments, we shall need some auxiliary segments that are not an intrinsic part of the tree but will be used in representing the discrete tree as a continuous flow. Suppose a vertex  $i$  in  $n$ -th generation has no children. Let  $j$  be the maximal vertex in generation  $n + 1$  among those having their parents preceding  $i$  in generation  $n$ . Then an auxiliary segment of type I connects the points  $(n, i - 1)$  and  $(n + 1, j - 1)$ . If vertex 1 in  $n$ -th generation has no children, points  $(n, 0)$  and  $(n + 1, 0)$  are connected by an auxiliary segment of type II. Auxiliary segments of type III connect points  $(n, i - 1)$  and  $(n, i)$  for  $1 \leq i \leq X_n - 1$ .

Every bounded connected component of the complement to the union of the above segments on the plane is either a parallelogram with two vertical sides of length 1, or a triangle with one vertical side of length 1. One can treat both shapes as trapezoids (with one of the parallel sides having zero length in the case of triangle).

For each trapezoid, we shall establish a bijection with the unit square and define the monotone flow to act along the images of the “horizontal” segments of the square. A graphic illustration of the construction is given on Figure 1, and we proceed to describe it precisely.

Each trapezoid  $L$  of this family has vertices  $g_{0,0} = (n, i_{0,0})$ ,  $g_{0,1} = (n, i_{0,1})$ ,  $g_{1,0} = (n+1, i_{1,0})$ ,  $g_{1,1} = (n+1, i_{1,1})$ , where  $i_{0,1} - i_{0,0} \in \{0, 1\}$  and  $i_{1,1} - i_{1,0} \in \{0, 1\}$ .

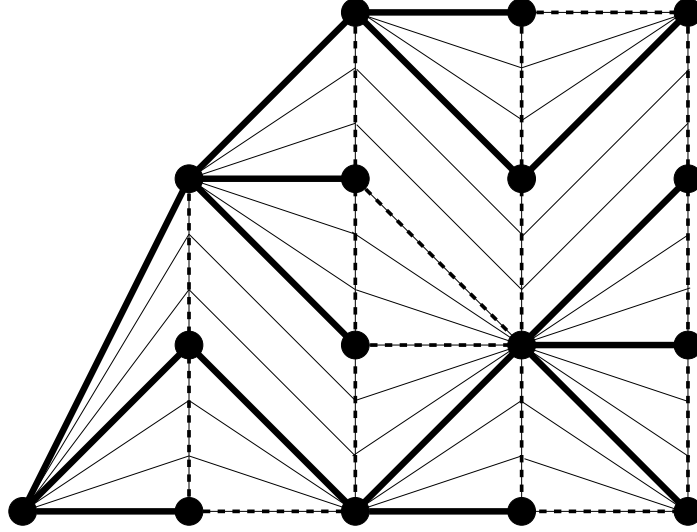


Figure 1: Construction of the continuous monotone flow.

Then, for every  $\alpha \in (0, 1)$  we define

$$g_m^L(\alpha) = g_{m,0} + \alpha(g_{m,1} - g_{m,0}), \quad m = 0, 1,$$

and

$$g^L(\alpha, s) = g_0^L(\alpha) + s(g_1^L(\alpha) - g_0^L(\alpha)), \quad s \in (0, 1).$$

This definition introduces a coordinate system in  $L$ , i.e., a bijection between  $L$  and the unit square  $(0, 1)^2$ . We are going to use it to construct the monotone map associated with the tree for times  $t_0, t_1$  assuming that there is  $n \in \{0\} \cup \mathbb{N}$  such that  $n < t_0 \leq t_1 < n + 1$ . Let us take any  $x$  such that  $(x, t_0)$  belongs to one of the trapezoids  $L$ . Then there is a unique number  $\alpha(x, t_0) \in (0, 1)$  such that  $g^L(\alpha(x, t_0), \{t_0\}) = x$ , where  $\{\cdot\}$  denotes the fractional part. We can define  $g^{t_0, t_1}(x) = g^L(\alpha(x, t_0), \{t_1\})$ . This strictly increasing function can be consistently and uniquely extended by continuity to points  $x$  such that  $(x, t_0)$  belongs either to a regular segment in the tree representation or an auxiliary segment of type I or II. This function  $g^{t_0, t_1}$  also uniquely defines a monotone graph  $\tilde{\Gamma}^{t_0, t_1} = \tilde{\Gamma}^{t_0, t_1}(\tau)$  depending continuously on  $t_0, t_1$ . Next, if we allow  $t_0$  and  $t_1$  to take values  $n$  and  $n + 1$ , then we can construct the associated monotone graph as the limit in  $(\mathbb{M}, \rho)$  of the monotone graphs associated to the increasing functions defined above (as  $t_0 \rightarrow n$  or  $t_1 \rightarrow n + 1$ ). Notice that the resulting monotone graphs may have intervals of constancy and shocks (i.e., contain horizontal and vertical

segments). Now we can take any  $(t_0, t_1) \in \Delta_\infty$  and define

$$\tilde{\Gamma}^{t_0, t_1} = \tilde{\Gamma}^{[t_1], t_1} \circ \tilde{\Gamma}^{[t_1]-1, t_1} \circ \dots \circ \tilde{\Gamma}^{[t_0]+1, [t_0]+2} \circ \tilde{\Gamma}^{t_0, [t_0]+1},$$

which results in a continuous monotone flow  $(\tilde{\Gamma}^{t_0, t_1}(\tau))_{(t_0, t_1) \in \Delta_\infty}$ .

To state our main result we need to introduce a rescaling of this family. For every  $n \in \mathbb{N}$ , we define

$$\Gamma_n^{t_0, t_1}(\tau) = \left\{ \left( \frac{x}{\mu n}, \frac{y}{\mu n} \right) : (x, y) \in \tilde{\Gamma}^{nt_0, nt_1}(\tau) \right\}, \quad (t_0, t_1) \in \Delta_\infty. \quad (14)$$

For each  $T > 0$  we can consider the uniform distance  $\rho_T$  on monotone flows in  $\mathbb{M}[0, T]$  introduced in (11), and define the locally uniform (LU) metric on  $\mathbb{M}[0, \infty)$  via

$$d(\Gamma_1, \Gamma_2) = \sum_{m=1}^{\infty} 2^{-m} (\rho_m(\Gamma_1, \Gamma_2) \wedge 1).$$

**Theorem 6** *As  $n \rightarrow \infty$ , the random field  $(\Gamma_n^{t_0, t_1}(\tau))_{(t_0, t_1) \in \Delta}$  converges in distribution in LU metric to the stochastic flow of monotone graphs solving equation (9).*

The rest of the section is devoted to the proof of this theorem. First, it is sufficient to prove the convergence in distribution in  $\rho_m$  for all  $m$ . We take  $m = 1$  without loss of generality. Due to the classical Prokhorov theorem it is sufficient to demonstrate two facts:

1. The sequence of distributions of  $\Gamma_n(\tau)$  is tight w.r.t.  $\rho_1$ .
2. Any limit point for the sequence of distributions coincides with the distribution of the flow generated by SPDE.

Subsection 8.2 is devoted to the proof of the first statement and subsection 8.3 to the second one.

## 8.2 Tightness

We have to show that for any  $\varepsilon > 0$  there is a set  $K \subset \mathbb{M}[0, 1]$  that is compact in  $d_1$  and  $\mathbb{P}\{\Gamma_n \in K\} > 1 - \varepsilon$  for all  $n$  (from now on we do not distinguish between  $\Gamma_n$  and its restriction on  $[0, 1]$ .)

The construction below depends on the values of constants  $\alpha > 0, \beta > \gamma > 0$  and  $b > 0$ . For  $m \in \mathbb{N}$ , consider a set

$$R_m = \left\{ (k2^{-m}, j2^{-m}) : 0 \leq k \leq 2^m; 0 \leq j \leq 2^{m(1+\alpha)} \right\},$$

and also define  $R = \bigcup_m R_m$ . For a function  $f$  we denote its  $\beta$ -Hölder constant by  $\mathcal{H}_\beta(f)$ :

$$\mathcal{H}_\beta(f) = \sup_{t_1, t_2: t_1 \neq t_2} \frac{|f(t_1) - f(t_2)|}{|t_1 - t_2|^\beta}.$$

Let  $K_m$  be the set of all monotone flows  $(\Gamma, Z, U)$  in  $\mathbb{M}[0, 1]$  such that  $\sup_t Z(t) \leq 2^{m\alpha}$ ,  $\mathcal{H}_\beta(Z) \leq b2^{m\gamma}$ , and for each  $(t_0, x) \in R_m$  satisfying  $x < Z(t_0)$ ,  $\mathcal{H}_\beta(U(x, t_0, \cdot)) \leq b2^{m\gamma}$ . Denote  $K = \bigcap_{m=m_0}^\infty K_m$ . Note that we suppress the dependence of  $K$  on  $b$  and  $m_0$ .

**Lemma 9** *For any  $b$  and  $m_0$ , the set  $K$  is precompact in  $\mathbb{M}[0, 1]$ .*

PROOF: For any sequence  $(\Gamma_n, Z_n, U_n)_{n \in \mathbb{N}}$  in  $K$ , we must show that it contains a convergent subsequence.

Using the Hölder continuity, the Arzelà–Ascoli compactness criterion, and the classical diagonal procedure, we can extract a subsequence  $n'$  such that for each  $m \in \mathbb{N}$  and each  $(t_0, x) \in R_m$ ,  $U_{n'}(x, t_0, \cdot)$  converges uniformly to a limiting trajectory  $U_\infty(x, t_0, \cdot)$  with  $\mathcal{H}_\beta(U_\infty(x, t_0, \cdot)) \leq b2^{m\gamma}$ , and  $Z_{n'}$  converges uniformly to a limiting trajectory  $Z_\infty$ .

We need to show that there is a unique continuous monotone flow  $\Gamma_\infty$  compatible with these trajectories and that  $\Gamma_{n'}$  converges uniformly to  $\Gamma_\infty$ . Let us construct  $\Gamma_\infty = (\Gamma_\infty, Z_\infty, U_\infty)$ . Notice that  $Z_\infty$  is already at our disposal. Take any point  $(t_0, x) \in [0, 1] \times \mathbb{R}_+$  such that  $x < Z_\infty(t_0)$ . For any  $m$ , we set

$$k = k(m) = \lfloor 2^m t_0 \rfloor,$$

$$j^- = j^-(m) = \lfloor 2^m (x - 2^{m\gamma} \cdot (2^{-m})^\beta) \rfloor,$$

and

$$j^+ = j^+(m) = \lfloor 2^m (x + 2^{m\gamma} \cdot (2^{-m})^\beta) \rfloor + 1.$$

Then,

$$j^+ 2^{-m} - j^- 2^{-m} < 2(2^{-m} + 2^{m\gamma} \cdot 2^{-m\beta}),$$

and

$$j^- 2^{-m} + 2^{m\gamma} \cdot (2^{-m})^\beta \leq x \leq j^+ 2^{-m} - 2^{m\gamma} \cdot (2^{-m})^\beta.$$

For sufficiently large  $m$ ,  $j^+ 2^{-m} < Z_\infty(k 2^{-m})$ , so that  $U_\infty(j^\pm 2^{-m}, k 2^{-m}, t_0)$  are already defined. Using the bound on the Hölder constant, we get

$$\begin{aligned} x - 2^{-m} - 2 \cdot 2^{m(\gamma-\beta)} &\leq U_\infty(j^- 2^{-m}, k 2^{-m}, t_0) \leq x \\ &\leq U_\infty(j^+ 2^{-m}, k 2^{-m}, t_0) \leq x + 2^{-m} + 2 \cdot 2^{m(\gamma-\beta)}. \end{aligned}$$

For any  $t_1 \geq t_0$ , we define

$$U^-(x, t_0, t_1) = \sup_m U_\infty(k(m)2^{-m}, j^-(m)2^{-m}, t_1),$$

and

$$U^+(x, t_0, t_1) = \inf_m U_\infty(k(m)2^{-m}, j^+(m)2^{-m}, t_1).$$

It is clear that  $U^-(x, t_0, t_1) \leq U^+(x, t_0, t_1)$  and if  $x < y$  then  $U^+(x, t_0, t_1) \leq U^-(y, t_0, t_1)$ . In particular, both quantities are monotone in first argument and there is a monotone graph  $\Gamma_\infty^{t_0, t_1}$  such that  $U^-(x, t_0, t_1)$  and  $U^+(x, t_0, t_1)$  are its left continuous and, respectively, right continuous representatives. Also any monotone function compatible with  $(U(x, t_0, t_1), (t_0, x) \in R)$ , must represent the same monotone graph.

The resulting family of monotone graphs  $(\Gamma_\infty^{t_0, t_1})$  is easily seen to satisfy all the properties in the definition of a monotone flow. Let us prove only Property 2, the continuity of  $\Gamma_\infty^{t_0, t_1}$  in  $t_0$  and  $t_1$ . It is sufficient to show that  $\Gamma_\infty^{t_0, t_1}$  converges to the graph of the identity map on  $[0, Z_\infty(t)]$  as  $t_0$  and  $t_1$  converge to  $t$  from below and above respectively. To see the latter, we imbed a small time interval  $(t_0, t_1)$  into some dyadic interval  $[k2^{-m}, (k+1)2^{-m}]$  with a large  $m$ . On  $[k2^{-m}, (k+1)2^{-m}]$ , the monotone flow  $\Gamma_\infty$  displaces all dyadic points with denominator  $2^{-m}$  by at most  $2^{m(\gamma-\beta)}$ . Since  $2^{m(\gamma-\beta)} \rightarrow 0$  as  $m \rightarrow \infty$ , the continuity follows.

Now we shall prove the convergence to this monotone flow  $\Gamma_\infty$ . For any  $\varepsilon > 0$ , we will show that for sufficiently large values of  $n'$ ,  $d_1(\Gamma_\infty, \Gamma_{n'}) < \varepsilon$ . First, for any  $m \in \mathbb{N}$ , we can find  $n_0(m)$  such that for all  $n' > n_0(m)$  and all  $(t_0, x) \in R_m$ ,  $\|U_n(x, t_0, \cdot) - U_\infty(x, t_0, \cdot)\| < \varepsilon/2$ , where  $\|\cdot\|$  denotes the sup-norm. For any  $(t_0, x)$  choose points  $k, j^-, j^+$  as above and denote

$$\begin{aligned} y_{n'}^\pm &= U_{n'}(j^\pm 2^{-m}, k2^{-m}, t_0), \\ y_\infty^\pm &= U_\infty(j^\pm 2^{-m}, k2^{-m}, t_0). \end{aligned}$$

Then

$$\begin{aligned} x - 2^{-m} - 2 \cdot 2^{m(\gamma-\beta)} &\leq y_{n'}^- \leq x \leq y_{n'}^+ \leq x + 2^{-m} + 2 \cdot 2^{m(\gamma-\beta)}, \\ x - 2^{-m} - 2 \cdot 2^{m(\gamma-\beta)} &\leq y_\infty^- \leq x \leq y_\infty^+ \leq x + 2^{-m} + 2 \cdot 2^{m(\gamma-\beta)}. \end{aligned}$$

The image of  $x$  under  $\Gamma_{n'}^{t_0, t_1}$  (viewed as a multivalued map) for every  $t_1 > t_0$  is contained in the image of  $[y_{n'}^-, y_{n'}^+]$  under  $\Gamma_{n'}^{t_0, t_1}$ . Therefore, the definition of  $y_{n'}^\pm$  and  $y_\infty^\pm$  implies that if  $n' > n_0(m)$ , then any point in the image of

$[y_{n'}^-, y_{n'}^+]$  under  $\Gamma_{n'}^{t_0, t_1}$  is at most at  $\varepsilon/2$  from the image of  $[y_\infty^-, y_\infty^+]$  under  $\Gamma_\infty^{t_0, t_1}$ . On the plane, the distance between any point  $(x, x') \in \Gamma_{n'}^{t_0, t_1}$  and the graph  $\Gamma_\infty^{t_0, t_1}$  restricted to  $[y_\infty^-, y_\infty^+]$  does not exceed  $2^{-m} - 2 \cdot 2^{m(\gamma-\beta)} + \varepsilon/2$ . So it is sufficient to choose  $m$  so that  $2^{-m} - 2 \cdot 2^{m(\gamma-\beta)} < \varepsilon/2$  to have that distance bounded by  $\varepsilon$ . Similarly, the same bound holds true for the distance between any point  $(x, x') \in \Gamma_\infty^{t_0, t_1}$  and the restriction of  $\Gamma_{n'}^{t_0, t_1}$  to  $[y_\infty^-, y_\infty^+]$ . We conclude that for the above choice of  $m$  and for any  $n' > n_0(m)$ , the uniform distance between  $\Gamma_{n'}$  and  $\Gamma_\infty$  is bounded by  $\varepsilon$ , and the convergence implying the precompactness of  $K$  follows.  $\square$

Our next goal is to show that for any  $\varepsilon > 0$ , the numbers  $\alpha, \beta, \gamma, b$  can be chosen so that  $P\{\Gamma_n \in K\} > 1 - \varepsilon$  for all  $n$ .

**Lemma 10** *Suppose there are positive numbers  $C, r$  such that for all  $m \in \mathbb{N}$ , all  $(t_0, x) \in R_m$ , a continuous process  $U(x, t_0, \cdot)$  satisfies*

$$E|U(x, t_0, t) - U(x, t_0, t')|^r \leq C|t - t'|^{r/2}, \quad t_0 \leq t \leq t'. \quad (15)$$

If

$$r/2 - 1 > r\beta, \quad (16)$$

then there is a constant  $C_1(C, \beta, r)$  such that for all  $m \in \mathbb{N}$  and any  $b > 0$ ,

$$P\left\{\sup_{(t_0, x) \in R_m} \mathcal{H}_\beta(U(x, t_0, \cdot)) > b2^{\gamma m}\right\} \leq C_1(C, \beta, r) \frac{2^{(2+\alpha-r\gamma)m}}{b^r}.$$

PROOF: Let us fix  $x, t_0$  and (for brevity) denote  $U(t) = U(x, t_0, t)$ ,

Let us estimate  $P\{\mathcal{H}_\beta(U) \geq c\}$  for a  $c \geq 2$ . If  $\mathcal{H}_\beta(U) \geq c$ , then there are times  $t < t'$  such that

$$|U(t') - U(t)| > c(t' - t)^\beta.$$

We can find  $m$  such that  $2^{-m} \leq t' - t < 2^{-m+1}$ . Then

$$|U(t') - U(t)| > c2^{-m\beta}. \quad (17)$$

There is an integer  $j$  such that  $j2^{-m} \in [t, t']$ . We can find sequences  $(\kappa_k)_{k=m}^\infty$  and  $(\kappa'_k)_{k=m}^\infty$  such that each  $\kappa_k$  and  $\kappa'_k$  is 0 or 1 and

$$t' - j2^{-m} = \sum_{k=m}^\infty \kappa'_k 2^{-k},$$

$$j2^{-m} - t = \sum_{k=m}^{\infty} \kappa_k 2^{-k}.$$

Assuming that on each respective dyadic interval of length  $2^{-k}$ , the increment of  $U$  does not exceed  $(1 - 2^{-\beta})c2^{-\beta k-1}$  in absolute value, we have

$$\begin{aligned} |U(t') - U(t)| &\leq (1 - 2^{-\beta})c2^{-1} \left( \sum_{k=m}^{\infty} \kappa'_k 2^{-\beta k} + \sum_{k=m}^{\infty} \kappa_k 2^{-\beta k} \right) \\ &\leq (1 - 2^{-\beta})c2^{-1} \cdot 2 \sum_{k=m}^{\infty} 2^{-\beta k} \\ &\leq c2^{-m\beta}, \end{aligned}$$

which contradicts (17). Therefore, denoting  $\nu(\beta) = (1 - 2^{-\beta})2^{-1}$ , we see that for some  $k$  there is a dyadic interval  $[l2^{-k}, (l+1)2^{-k}]$  of length  $2^{-k}$  such that

$$|U(l2^{-k}) - U((l+1)2^{-k})| > \nu(\beta)c2^{-\beta k}.$$

We use Markov's inequality and condition (15) to see that an upper bound for the probability of this event is:

$$\begin{aligned} &\sum_{k=0}^{\infty} 2^k \mathbf{P} \left\{ |U(l2^{-k}) - U((l+1)2^{-k})| > \nu(\beta)c2^{-\beta k} \right\} \\ &\leq \frac{1}{c^r \nu(\beta)^r} \sum_{k=0}^{\infty} 2^k \mathbf{E} |U(l2^{-k}) - U((l+1)2^{-k})|^r 2^{r\beta k} \\ &\leq \frac{C}{c^r \nu(\beta)^r} \sum_{k=0}^{\infty} 2^k 2^{-kr/2} 2^{r\beta k} \leq \frac{C_1(C, \beta, r)}{c^r}, \end{aligned}$$

where

$$C_1(C, \beta, r) = \frac{C}{\nu(\beta)^r} \sum_{k=0}^{\infty} 2^k 2^{-kr/2} 2^{r\beta k}.$$

Notice that  $C_1(C, \beta, r) < \infty$  due to (16). The lemma follows as we set  $c = b2^{\gamma m}$  since there are  $2^{(2+\alpha)m}$  points in  $R_m$ .  $\square$

The proof of the following lemma is similar to that of Lemma 10, and we omit it.

**Lemma 11** *If  $r/2 - 1 > r\beta$  and  $Z$  is a continuous process satisfying*

$$\mathbf{E} |Z(t') - Z(t)|^r \leq C |t' - t|^{r/2}, \quad t, t' \in [0, 1], \quad (18)$$

then

$$\mathbb{P}\{\mathcal{H}_\beta(Z) \geq b2^{\gamma m}\} \leq C_1(C, \beta, r) \frac{2^{-r\gamma m}}{b^r}.$$

**Lemma 12** *Let*

$$2 + \alpha < r\gamma. \quad (19)$$

*Suppose  $\Gamma = (\Gamma, Z, U)$  is a monotone flow such that processes  $U$  satisfy the conditions of Lemma 10, and process  $Z$  satisfies the conditions of Lemma 11. Suppose there is a constant  $\bar{C}$  such that*

$$\mathbb{E} \sup_{t \in [0,1]} Z(t) \leq \bar{C}. \quad (20)$$

*then there are  $b = b(\alpha, \beta, \gamma, r, C, \bar{C})$  and  $m_0 = m_0(\alpha, \beta, \gamma, r, C, \bar{C})$  such that*

$$\mathbb{P}\{\Gamma \in K\} > 1 - \varepsilon.$$

PROOF:

$$\begin{aligned} \mathbb{P}\{\Gamma \in K\} &\leq \sum_{m=1}^{\infty} \mathbb{P}\left\{ \sup_{(t_0, x) \in R_m} \mathcal{H}_\beta(U(x, t_0, \cdot)) > b2^{\gamma m} \right\} \\ &\quad + \sum_{m=1}^{\infty} \mathbb{P}\{\mathcal{H}_\beta(Z) \geq b2^{\gamma m}\} \\ &\quad + \sum_{m=m_0}^{\infty} \mathbb{P}\left\{ \sup_{t \in [0,1]} Z(t) > 2^{\alpha m} \right\}. \quad (21) \end{aligned}$$

Lemma 10 and condition (19) imply that for sufficiently large  $b$ , the first term in the r.h.s. is less than  $\varepsilon/3$ . It also follows from Lemma 11 that for sufficiently large  $b$ , the second term in the r.h.s. is less than  $\varepsilon/3$ . Finally, we can use (20) and Markov's inequality to choose  $m_0$  so that the last term in the r.h.s. is less than  $\varepsilon/3$  completing the proof.  $\square$

Since we want (16) and (19) to be satisfied along with  $\gamma < \beta$ , we need to have

$$2 + \alpha < r\beta < \frac{r}{2} - 1.$$

This can be satisfied if we take  $r = 8$ ,  $1/4 < \gamma < \beta < 3/8$ ,  $0 < \alpha < 8\gamma - 2$ , and our goal is to show that all other conditions are satisfied with this choice of parameters for the sequence of random monotone flows  $\Gamma_n = (\Gamma_n, Z_n, U_n)$  defined in (14).

From now on we assume that  $U_n$  is the special right-continuous trajectory representation of  $\Gamma_n$ , see (12).



**Lemma 13** *For all  $n$ , the process  $Z_n$  restricted to times  $0, n^{-1}, 2n^{-1}, 3n^{-1}, \dots$  is a submartingale with respect to its natural filtration. For any  $r \in \mathbb{N}$ ,  $\mathbb{E}Z_n^r(t)$  is uniformly bounded in  $t \in [0, 1]$  and  $n \in \mathbb{N}$ .*

PROOF: It is sufficient to consider the process  $X_n$  counting vertices in the  $n$ -th generation of the discrete random tree and show that it is a submartingale and that for every  $r$  there is a constant  $c_r$  such that for all  $n \in \mathbb{N}$ ,

$$\mathbb{E}X_n^r < c_r n^r. \quad (22)$$

The latter follows from the following one: for every  $r$  there is a polynomial  $P_{r-1}$  of degree  $r-1$  such that

$$\mathbb{E}[X_{n+1}^r | X_n] \leq X_n^r + P_{r-1}(X_n). \quad (23)$$

Taking expectations of both sides of (23) and applying a straightforward induction procedure in  $r$  and  $n$ , we obtain the growth estimate (22).

To prove (23), we define  $P(r, k)$  to be the set of all partitions of  $r$  into  $k$  nonnegative integers:

$$P(r, k) = \{(j_1, \dots, j_k) : j_1 \geq j_2 \geq \dots \geq j_k \geq 0, j_1 + j_2 + \dots + j_k = r\},$$

and write

$$\begin{aligned} \mathbb{E}[X_{n+1}^r | X_n = k] &= \frac{1}{k} \sum_{0 \leq i_1, \dots, i_k \leq D} (i_1 + \dots + i_k)^{r+1} p_{i_1} \dots p_{i_k} \\ &= \frac{1}{k} \sum_{0 \leq i_1, \dots, i_k \leq D} \sum_j \binom{r+1}{j_1, \dots, j_k} i_1^{j_1} i_2^{j_2} \dots i_k^{j_k} p_{i_1} \dots p_{i_k} \\ &= \frac{1}{k} \sum_{0 \leq i_1, \dots, i_k \leq D} \sum_{j \in P(r+1, k)} K_j(r) i_1^{j_1} i_2^{j_2} \dots i_k^{j_k} p_{i_1} \dots p_{i_k} \\ &= \frac{1}{k} \sum_{j \in P(r+1, k)} K_j(r) B_{j_1} B_{j_2} \dots B_{j_k}. \end{aligned} \quad (24)$$

Here  $i_1, \dots, i_k$  denote the number of children of each of  $k$  vertices of generation  $n$ , and in the third line we used the symmetry of the expression and grouped together monomials producing the same partition  $j \in P(r+1, k)$  under the monotone rearrangement of their degrees (we agree that  $0^0 = 1$ ). The numbers  $B_m$  have been introduced in (2). The positive constants  $K_j(r)$  satisfy

$$\sum_{j \in P(r+1, k)} K_j(r) = k^{r+1}. \quad (25)$$

Notice that if  $k \geq r + 1$  then

$$K_{1,1,\dots,1,0,0,\dots,0} = k(k-1) \dots (k-r) = k^{r+1} - \tilde{P}_r(k),$$

where  $\tilde{P}_r(k)$  is a polynomial of degree at most  $r$ . In this case, due to (25), the sum of all other constants equals  $\tilde{P}_r(k)$ . Now (23) follows straight from (24).

The fact that  $(X_n)$  is a submartingale follows from

$$\mathbb{E}[X_{n+1} - X_n | \sigma(X_0, \dots, X_n)] = \mu,$$

which was established in [Bak]. This identity can also be derived from the computation in the proof of Lemma 14.  $\square$

**Lemma 14** *The conditions of Lemma 10 are satisfied for  $U_n$  uniformly in  $n$ , if we choose  $r = 8$ ,  $1/4 < \gamma < \beta < 3/8$ ,  $0 < \alpha < 8\gamma - 2$ .*

PROOF: We have to prove that (15) is satisfied for  $U_n(x, t_0, \cdot)$  uniformly in  $n, x, t_0, t, t'$ . Suppose first that  $(t_0, x)$  is a grid point, i.e.,  $t_0 n$  and  $x\mu n$  are integers. Denote by  $V_j$  the size of progeny in generation  $j$  generated by first  $x\mu n + 1$  vertices in  $t_0 n$ -th generation. Then  $U_n(x, t_0, j/n) = (V_j - 1)/(\mu n)$ . Let us compute

$$\begin{aligned} & \mathbb{E} \left[ \frac{V_{m+1}}{\mu n} \middle| \frac{V_m}{\mu n} = \frac{l}{\mu n}, \frac{X_m}{\mu n} = \frac{k}{\mu n} \right] \\ &= \frac{1}{\mu n k} \sum_{0 \leq i_1, \dots, i_k \leq D} (i_1 + \dots + i_l)(i_1 + \dots + i_k) p_{i_1} \dots p_{i_k} \\ &= \frac{1}{\mu n k} \left[ l \sum_{i_1} i_1^2 p_{i_1} \dots p_{i_k} + l(k-1) \sum_{i_1 \neq i_2} i_1 i_2 p_{i_1} \dots p_{i_k} \right] \\ &= \frac{1}{\mu n k} [lB_2 + l(k-1)] \\ &= \frac{l}{\mu n} \left( 1 + \frac{\mu}{k} \right). \end{aligned}$$

Therefore,

$$\mathbb{E} \left[ \frac{V_{m+1}}{\mu n} - \frac{V_m}{\mu n} \middle| \frac{V_m}{\mu n} = \frac{l}{\mu n}, \frac{X_m}{\mu n} = \frac{k}{\mu n} \right] = \frac{1}{n} \frac{l}{k}. \quad (26)$$

Since  $0 \leq l/k \leq 1$ , the process  $\bar{A}_n(t)$  interpolating between the values of

$$A_n(t) = \frac{1}{\mu n} \sum_{j=0}^{[nt]-1} \frac{V_j}{X_j}$$

is a 1-Lipschitz process, and

$$B_n(t) = U_n(x, t_0, t) - \bar{A}_n(t)$$

restricted to  $t = mn^{-1}$ ,  $m = 0, 1, \dots, n$ , is a martingale. Since  $\bar{A}$  is 1-Lipschitz, we now need to prove that  $B_n$  satisfies a moment estimate analogous to (15) with constant in the r.h.s. independent of  $n$ . If  $t = mn^{-1}$  and  $t' = m'n^{-1}$ , then due to Burkholder's inequality (see [Shi96, Section VII.3]) and Minkowski's inequality (see [Shi96, II.6]), there is an absolute constant  $C_2$  such that

$$\begin{aligned} \mathbb{E}|B_n(t') - B_n(t)|^8 &\leq C_2 \mathbb{E} \left( \sum_{j=m}^{m'-1} (B_n((j+1)n^{-1}) - B_n(jn^{-1}))^2 \right)^4 \\ &\leq C_2 \left\| \sum_{j=m}^{m'-1} (B_n((j+1)n^{-1}) - B_n(jn^{-1}))^2 \right\|_4^4 \\ &\leq C_2 \left( \sum_{j=m}^{m'-1} \left\| (B_n((j+1)n^{-1}) - B_n(jn^{-1}))^2 \right\|_4 \right)^4 \\ &\leq C_2 (m' - m)^4 \max_{j < m'} \left\| (B_n((j+1)n^{-1}) - B_n(jn^{-1}))^2 \right\|_4^4 \\ &\leq C_2 n^4 (t' - t)^4 \max_{j < n} \mathbb{E} |B_n((j+1)n^{-1}) - B_n(jn^{-1})|^8. \end{aligned}$$

To estimate the r.h.s., we use the following statement:

**Lemma 15** *There is an absolute constant  $C_3$  such that*

$$\mathbb{E} |B_n((j+1)n^{-1}) - B_n(jn^{-1})|^8 \leq \frac{C_3}{n^4}.$$

The desired moment estimate follows directly:

$$\mathbb{E}|B_n(t') - B_n(t)|^8 \leq C_2 C_3 (t' - t)^4.$$

PROOF OF LEMMA 15:

$$\left| B_n \left( \frac{j+1}{n} \right) - B_n \left( \frac{j}{n} \right) \right| \leq \frac{1}{\mu n} + \left| \frac{V_{j+1}}{\mu n} - \frac{V_j}{\mu n} \right|,$$

so that, by convexity,

$$\mathbb{E} \left| B_n \left( \frac{j+1}{n} \right) - B_n \left( \frac{j}{n} \right) \right|^8 \leq 2^7 \frac{1}{(\mu n)^8} + 2^7 \mathbb{E} \left| \frac{V_{j+1}}{\mu n} - \frac{V_j}{\mu n} \right|^8,$$

and the lemma follows directly from the next result:

**Lemma 16** *For any even  $r \geq 2$ , there is a constant  $c'_r$  such that for all  $n$  and all  $j$ ,*

$$\mathbb{E} \left( \frac{V_{j+1}}{n} - \frac{V_j}{n} \right)^r \leq c'_r \frac{j^{r/2}}{n^r}. \quad (27)$$

PROOF: Let us estimate  $\mathbb{E}(V_{j+1} - V_j)^r$ .

$$\begin{aligned} & \mathbb{E}[(V_{j+1} - V_j)^r | V_j = l, X_j = k] \\ &= \frac{1}{k} \sum_{i_1, \dots, i_k} (i_1 + \dots + i_l - l)^r (i_1 + \dots + i_k - k) p_{i_1} \dots p_{i_k} \\ & \quad + \sum_{i_1, \dots, i_k} (i_1 + \dots + i_l - l)^r p_{i_1} \dots p_{i_k} \\ &= \frac{1}{k} \sum_{i_1, \dots, i_k} ((i_1 - 1) + \dots + (i_l - 1))^r ((i_1 - 1) + \dots + (i_k - 1)) p_{i_1} \dots p_{i_k} \\ & \quad + \sum_{i_1, \dots, i_k} ((i_1 - 1) + \dots + (i_l - 1))^r p_{i_1} \dots p_{i_k} \\ &= \frac{1}{k} \sum_{i_1, \dots, i_k} ((i_1 - 1) + \dots + (i_l - 1))^{r+1} p_{i_1} \dots p_{i_k} \\ & \quad + \frac{1}{k} \sum_{i_1, \dots, i_k} ((i_1 - 1) + \dots + (i_l - 1))^r ((i_{l+1} - 1) + \dots + (i_k - 1)) p_{i_1} \dots p_{i_k} \\ & \quad + \sum_{i_1, \dots, i_k} ((i_1 - 1) + \dots + (i_l - 1))^r p_{i_1} \dots p_{i_k} \\ &= S_1 + S_2 + S_3 \end{aligned}$$

We begin with  $S_3$ . Since

$$\sum_{i_{l+1}, \dots, i_k} p_{i_{l+1}} \dots p_{i_k} = 1,$$

$$\begin{aligned} S_3 &= \sum_{j \in P(r, l)} \sum_{i_1, \dots, i_l} (i_1 - 1)^{j_1} \dots (i_l - 1)^{j_l} p_1 \dots p_l \\ &= \sum_{j \in P(r, l)} K_j \bar{B}_{j_1} \dots \bar{B}_{j_l}, \end{aligned}$$

where all monomials with similar degree  $j$  are grouped together, and

$$\bar{B}_j = \sum_i (i - 1)^j p_i.$$

Notice that  $\bar{B}_1 = 0$ , so that if one of the terms in  $j_1 + \dots + j_l = r$  equals 1, the contribution from that term to  $S_3$  is 0. In other words, only partitions of  $r$  containing no 1's contribute to  $S_3$ . Each of these partitions contains at most  $r/2$  nonzero components, and therefore the associated coefficient  $K_j$  (the number of monomials associated to this partition) is bounded by a degree  $r/2$  polynomial in  $l$ . The same analysis shows that  $S_2 = 0$ , and  $S_1$  is bounded by a polynomial of degree  $r/2$ . We conclude that there is a constant  $C'_r$  (independent of  $k$ ) such that

$$\mathbb{E}[(V_{j+1} - V_j)^r | V_j = l, X_j = k] \leq C'_r l^{r/2}. \quad (28)$$

The estimate (22) implies now that

$$\mathbb{E}(V_{j+1} - V_j)^r \leq C'_r \mathbb{E}X_j^{r/2} \leq c'_r j^{r/2},$$

so that (27) holds true.  $\square$

We assume now that the desired moment estimate for  $(t_0, x)$  on the grid holds true with a constant  $C$ . We need is to estimate  $\mathbb{E}|U_n(x, t_0, t') - U_n(x, t_0, t)|^8$  for arbitrary  $n, x, t_0, t, t'$ .

Let  $j = [tn]$ ,  $j' = [t'n]$  and  $N = j' - j$ . We shall assume that  $t_0 n < j$ ; the proof can be easily modified to treat the opposite situation.

We are going to consider three cases: (i)  $N = 0$ , (ii)  $N = 1$ , (iii)  $N \geq 2$ .

In case (i), the evolution of  $U_n(x, t_0, t)$  is linear in  $t \in [jn^{-1}, (j+1)n^{-1}]$ , so that, comparing it to the evolution along the regular edges of the tree imbedding, we see that

$$\begin{aligned} \mathbb{E} \left( \frac{U_n(x, t_0, t') - U_n(x, t_0, t)}{t' - t} \right)^8 &= \mathbb{E} \left( \frac{U_n(x, t_0, \frac{j+1}{n}) - U_n(x, t_0, \frac{j}{n})}{1/n} \right)^8 \\ &\leq n^8 \sum_l \mathbb{E} \left[ (U_n(x, t_0, \frac{j+1}{n}) - U_n(x, t_0, \frac{j}{n}))^8 | U_n(x, t_0, \frac{j}{n}) \in (\frac{l}{\mu n}, \frac{l+1}{\mu n}] \right], \\ &\leq \sum_l \left( \mathbb{E}[(V_{j+1} - V_j)^8 | V_j = l+1] \mathbb{P} \left\{ U_n(x, t_0, \frac{j}{n}) \in (\frac{l}{\mu n}, \frac{l+1}{\mu n}] \right\} \right. \\ &\quad \left. + \mathbb{E}[(V_{j+1} - V_j)^8 | V_j = l] \mathbb{P} \left\{ U_n(x, t_0, \frac{j}{n}) \in (\frac{l}{\mu n}, \frac{l+1}{\mu n}] \right\} \right). \end{aligned}$$

Inequality (28) implies now that we can continue the estimate as

$$\begin{aligned} \mathbb{E}(U_n(x, t_0, t') - U_n(x, t_0, t))^8 &\leq 2(t' - t)^8 C'_8 \sum_l (l+1)^4 \mathbb{P} \left\{ U_n(x, t_0, \frac{j}{n}) \in (\frac{l}{\mu n}, \frac{l+1}{\mu n}] \right\} \\ &\leq 2(t' - t)^8 \mu^4 n^4 C'_8 \mathbb{E} \left( U_n(x, t_0, \frac{j}{n}) + \frac{1}{\mu n} \right)^4 \\ &\leq 2(t' - t)^4 \mu^4 C'_8 C_4^*, \end{aligned} \quad (29)$$

where

$$C_4^* = \sup_{n \in \mathbb{N}} \sup_{t \in [0,1]} \mathbb{E} \left( Z_n(t) + \frac{1}{\mu n} \right)^4,$$

and in the last line we used the fact that  $t' - t \leq n^{-1}$ . In case (ii),

$$\begin{aligned} & \mathbb{E} \left( \frac{U_n(x, t_0, t') - U_n(x, t_0, t)}{t' - t} \right)^8 \\ & \leq \mathbb{E} \left( \frac{U_n(x, t_0, \frac{j+1}{n}) - U_n(x, t_0, t)}{\frac{j+1}{n} - t} \right)^8 + \mathbb{E} \left( \frac{U_n(x, t_0, t') - U_n(x, t_0, \frac{j+1}{n})}{t' - \frac{j+1}{n}} \right)^8 \\ & \leq \mathbb{E} \left( \frac{U_n(x, t_0, \frac{j+1}{n}) - U_n(x, t_0, \frac{j}{n})}{1/n} \right)^8 + \mathbb{E} \left( \frac{U_n(x, t_0, \frac{j+2}{n}) - U_n(x, t_0, \frac{j+1}{n})}{1/n} \right)^8, \end{aligned}$$

and for each of the terms in the r.h.s. we can proceed as in case (i) and obtain an estimate analogous to (29).

Case (iii). Let us estimate  $\mathbb{E}(U_n(x, t_0, t') - U_n(x, t_0, t))_+^8$  first. Let  $\xi = [U(x, t_0, t)\mu n]$ . Suppose  $U(x, t_0, t') \geq U(x, t_0, t)$ . Then the monotonicity of the flow implies that at least one of the following two conditions holds true:

1.  $U(\frac{\xi}{\mu n}, \frac{j}{n}, \frac{j+1}{n}) - U(x, t_0, t) \geq \frac{1}{2}(U(x, t_0, t') - U(x, t_0, t)).$
2. For  $\eta = [U(x, t_0, \frac{j+1}{n})\mu n] + 1$ ,

$$U(\frac{\eta}{\mu n}, \frac{j+1}{n}, \frac{j'-1}{n}) \wedge U(\frac{\eta}{\mu n}, \frac{j+1}{n}, \frac{j'}{n}) - \frac{\eta}{\mu n} \geq \frac{1}{2}(U(x, t_0, t') - U(x, t_0, t)) - \frac{1}{\mu n}.$$

Consequently,

$$(U(x, t_0, t') - U(x, t_0, t))_+ \leq 2(R_1 \wedge R_2 \wedge R_3),$$

where

$$\begin{aligned} R_1 &= U \left( \frac{\xi}{\mu n}, \frac{j}{n}, \frac{j+1}{n} \right) - U(x, t_0, t), \\ R_2 &= U \left( \frac{\eta}{\mu n}, \frac{j+1}{n}, \frac{j'-1}{n} \right) - \frac{\eta}{\mu n} + \frac{1}{n} \\ R_3 &= U \left( \frac{\eta}{\mu n}, \frac{j+1}{n}, \frac{j'}{n} \right) - \frac{\eta}{\mu n} + \frac{1}{n}. \end{aligned}$$

Therefore,

$$\mathbb{E}(U(x, t_0, t') - x)_+^8 \leq 2^8 (\mathbb{E}R_1^8 + \mathbb{E}R_2^8 + \mathbb{E}R_3^8).$$

Notice that the random variables  $R_1, R_2, R_3$  are defined in terms of the flow trajectories emitted from lattice points. Although these lattice points defined through  $\xi$  and  $\eta$  are random, they are measurable with respect to the filtration of the flow, and therefore the above estimates imply that there is a constant  $C$  such that

$$\mathbb{E}(U(x, t_0, t') - U(x, t_0, t))_+^8 \leq C \left( \frac{j' - j}{n} \right)^4 \leq 2^4 C (t' - t)^4.$$

We can estimate  $\mathbb{E}(U(x, t_0, t') - U(x, t_0, t))_-^8$  similarly, which completes the proof.

### 8.3 Characterization of limit points

We begin with two auxiliary lemmas.

**Lemma 17** *Let  $m_0 \in \mathbb{N}$ . For each  $n \in \mathbb{N}$ , take a random number  $V(n)$  of vertices in generation  $m_0 n$  measurable with respect to the history of the tree up to generation  $m_0 n$ . For  $m \geq m_0 n$  we denote by  $V_m^{(n)}$  the total size of the progeny of these vertices in generation  $m$ . (In particular,  $V_{m_0 n}^{(n)} = V(n)$ .) If*

$$\mathbb{P}\{V(n)/n \geq n^{-\gamma}\} \rightarrow 0, \quad n \rightarrow \infty,$$

*for some constant  $\gamma \in (0, 1)$ , independent of  $n$ , then, for any  $m_1 > m_0$ ,*

$$\sup_{m_0 n \leq m \leq m_1 n} \frac{V_m^{(n)}}{n} \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

PROOF: Let  $X_m, m \geq 0$  is the process of total population sizes. If we show that

$$\sup_{m_0 n \leq m \leq m_1 n} \frac{V_m^{(n)}}{X_m} \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty, \tag{30}$$

then the lemma will follow since we can write

$$\sup_{m_0 n \leq m \leq m_1 n} \frac{V_m^{(n)}}{n} = \sup_{m_0 n \leq m \leq m_1 n} \frac{V_m^{(n)}}{X_m} \cdot \sup_{m_0 n \leq m \leq m_1 n} \frac{X_m}{n},$$

and use the convergence in distribution of  $\sup_{m_0 n \leq m \leq m_1 n} \frac{X_m}{n}$  to the maximum of a diffusion process.

To prove (30), we define a sequence of events

$$A_n = \{X_{m_0 n} \geq n^{1-\gamma/2}; V(n) \leq n^{1-\gamma}\},$$

and for any  $\varepsilon > 0$  write

$$\mathbf{P} \left\{ \sup_{m_0 n \leq m \leq m_1 n} \frac{V_m}{X_m} > \varepsilon \right\} \leq \mathbf{P}(A_n^c) + \mathbf{P} \left( \left\{ \sup_{m_0 n \leq m \leq m_1 n} \frac{V_m}{X_m} > \varepsilon \right\} \cap A_n \right).$$

Clearly,  $\mathbf{P}(A_n^c) \rightarrow 0$  as  $n \rightarrow \infty$  due to the assumptions of the lemma and the fact that the distribution of  $X_n/n$  converges weakly to a distribution with no atom at 0.

For the second term, we notice that  $\left(\frac{V_m^{(n)}}{X_m}\right)$  is a nonnegative bounded supermartingale w.r.t. filtration  $(\mathcal{F}_m^{(n)})$ , where for each  $m = m_0 n, \dots, m_1 n$ ,  $\mathcal{F}_m^{(n)}$  is the sigma-algebra generated by  $X_l, V_l, l = m_0 n, \dots, m$ . In fact,

$$\begin{aligned} & \mathbf{E} \left[ \frac{V_{m+1}^{(n)}}{X_{m+1}} \mid V_m^{(n)} = l, X_m = k \right] \\ &= \frac{1}{k} \sum_{\substack{0 \leq i_1, \dots, i_k \leq D \\ i_1 + \dots + i_k \neq 0}} \frac{i_1 + \dots + i_l}{i_1 + \dots + i_k} (i_1 + \dots + i_k) p_{i_1} \dots p_{i_k} \\ &= \frac{1}{k} \sum_{0 \leq i_1, \dots, i_k \leq D} (i_1 + \dots + i_l) p_{i_1} \dots p_{i_k} - \frac{1}{k} p_0^k \\ &= \frac{l}{k} - \frac{1}{k} p_0^k \\ &\leq \frac{l}{k}. \end{aligned}$$

Doob's maximal inequality for nonnegative supermartingales implies

$$\mathbf{P} \left( \left\{ \sup_{m_0 n \leq m \leq m_1 n} \frac{V_m}{X_m} > \varepsilon \right\} \cap A_n \right) \leq \frac{n^{1-\gamma}/n^{1-\gamma/2}}{\varepsilon} \rightarrow 0, \quad n \rightarrow \infty,$$

and (30) is proven.  $\square$

**Lemma 18** *Let  $\gamma \in (0, 1/2)$ . For any  $x > 0, t > 0$ , there is a constant  $C$  such that for all  $n \in \mathbb{N}$ ,*

$$\mathbf{P} \left\{ \left| U_n(x, t, \frac{[nt]+1}{n}) - \frac{[x\mu n]}{\mu n} \right| > n^{-\gamma} \right\} < C n^{-(1-2\gamma)}$$

PROOF: It is easily seen that for large  $n$ , due to the monotonicity of the flow, if

$$\left| U_n \left( x, t, \frac{[nt]+1}{n} \right) - \frac{[x\mu n]}{\mu n} \right| > n^{-\gamma},$$



then

$$U_n \left( \frac{[x\mu n]}{\mu n}, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} > 0,$$

and, moreover,

$$\left| U_n \left( \frac{[x\mu n]}{\mu n}, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} \right| > n^{-\gamma} - \frac{2}{n}.$$

Also, for large values of  $n$ , if

$$U_n \left( x, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} < -n^{-\gamma}.$$

then

$$U_n \left( \frac{[x\mu n]}{\mu n}, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} < 0$$

and, moreover,

$$\left| U_n \left( \frac{[x\mu n]}{\mu n}, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} \right| > n^{-\gamma} - \frac{2}{n}.$$

Therefore,

$$\begin{aligned} \mathbf{P} \left\{ \left| U_n \left( x, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} \right| > n^{-\gamma} \right\} \\ \leq \mathbf{P} \left\{ \left| U_n \left( \frac{[x\mu n]}{\mu n}, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} \right| > n^{-\gamma} - \frac{2}{n} \right\}, \end{aligned}$$

The moment estimate (28) from Lemma 16 implies that

$$\mathbf{E} \left| U_n \left( \frac{[x\mu n]}{\mu n}, t, \frac{[nt] + 1}{n} \right) - \frac{[x\mu n]}{\mu n} \right|^2 \leq C'_2 \frac{[x\mu n]}{\mu^2 n^2} \leq C'_2 \frac{x}{n},$$

and the desired estimate follows from Markov's inequality.  $\square$

Now we proceed to prove that any limiting point for the sequence of distributions of monotone flows  $\Gamma_n$  has to coincide with the monotone flow solving the SPDE (9) as discussed in Section 6.

Consider  $(x_{0,1}, t_0) = (0, 0)$ , a sequence of times  $0 < t_1 < t_2 < \dots < t_k$ , and for each  $i = 1, \dots, k$ , a sequence of nonnegative numbers

$$x_{i1} < \dots < x_{il(i)}.$$

For each  $i = 0, \dots, k$  and every  $j = 1, \dots, l(i)$  we define a family of processes

$$U_n^{i,j}(t) = \begin{cases} x_{ij}, & t < \frac{[nt_i]}{n}, \\ U_n\left(\frac{[x_{ij}\mu n]}{\mu n}, \frac{[nt_i]+1}{n}, t\right), & t \geq \frac{[nt_i]}{n}. \end{cases}$$

Due to Lemmas 17 and 18, it is sufficient to show that as  $n \rightarrow \infty$ , these processes jointly converge in distribution in sup-norm to a nonnegative diffusion process  $(U_\infty^{i,j})$  with drift

$$b^{ij}(y, t) = \frac{y_{ij}}{y_{01}} \mathbf{1}_{\{t > t_i; 0 \leq y_{ij} \leq y_{01}\}}, \quad i = 0, \dots, k, \quad j = 1, \dots, l(i),$$

and diffusion matrix

$$a^{i_1 j_1, i_2 j_2}(y, t) = (y_{i_1 j_1} \wedge y_{i_2 j_2}) \mathbf{1}_{\{t > t_{i_1} \vee t_{i_2}; 0 \leq y_{i_1 j_1}, y_{i_2 j_2} \leq y_{01}\}}, \\ i_1, i_2 = 0, \dots, k, \quad j_1 = 1, \dots, l(i_1), \quad j_2 = 1, \dots, l(i_2).$$

For any  $i$  and any  $y$ , these coefficients are constant on  $t \in [t_i, t_{i+1})$ , they are continuous and bounded in  $y$ , and define a well-posed martingale problem.

Let

$$\bar{U}_n^{ij}(t) = U_n^{ij}([nt]/n)$$

If  $t > t_i$  and  $U_n^{ij}(\frac{[nt]}{n}) \leq U_n^{01}(\frac{[nt]}{n})$ ,

$$\mathbb{E} \left[ U_n^{ij} \left( \frac{[nt]}{n} + \frac{1}{n} \right) \mid \mathcal{F}_{[nt]/n} \right] = U_n^{ij} \left( \frac{[nt]}{n} \right) + \frac{1}{n} \cdot \frac{U_n^{ij}(\frac{[nt]}{n})}{U_n^{01}(\frac{[nt]}{n})}.$$

So, we define

$$B_n^{ij}(t) = \frac{1}{n} \sum_{m=0}^{[nt]} \mathbf{1}_{\{m \geq nt_i; U_n^{ij}(\frac{[mt]}{n}) \leq U_n^{01}(\frac{[mt]}{n})\}} \frac{U_n^{ij}(\frac{m}{n})}{U_n^{01}(\frac{m}{n})} \quad (31)$$

and

$$M_n^{ij}(t) = \bar{U}_n^{ij}(t) - B_n^{ij}(t - \frac{1}{n}).$$

Next, a simple calculation based on the martingale property of  $M_n^{ij}$  and (31) shows that if  $t > t_i$ ,  $U_n^{i_1 j_1}(\frac{[nt]}{n}) \leq U_n^{01}(\frac{[nt]}{n})$ , and  $U_n^{i_2 j_2}(\frac{[nt]}{n}) \leq U_n^{01}(\frac{[nt]}{n})$  then

$$\begin{aligned} & \mathbb{E} \left[ M_n^{i_1 j_1} \left( \frac{[nt]}{n} + \frac{1}{n} \right) M_n^{i_2 j_2} \left( \frac{[nt]}{n} + \frac{1}{n} \right) - M_n^{i_1 j_1} \left( \frac{[nt]}{n} \right) M_n^{i_2 j_2} \left( \frac{[nt]}{n} \right) \mid \mathcal{F}_{[nt]/n} \right] \\ &= \mathbb{E} \left[ \bar{U}_n^{i_1 j_1} \left( \frac{[nt]}{n} + \frac{1}{n} \right) \bar{U}_n^{i_2 j_2} \left( \frac{[nt]}{n} + \frac{1}{n} \right) \mid \mathcal{F}_{[nt]/n} \right] \\ & \quad - \bar{U}_n^{i_1 j_1} \left( \frac{[nt]}{n} \right) \left( 1 + \frac{1}{n} \cdot \frac{B_2 - 1}{U_n^{01}(\frac{[nt]}{n})} \right) \bar{U}_n^{i_2 j_2} \left( \frac{[nt]}{n} \right) \left( 1 + \frac{1}{n} \cdot \frac{B_2 - 1}{U_n^{01}(\frac{[nt]}{n})} \right). \end{aligned} \quad (32)$$

Processes  $U_n^{i_1 j_1}$  and  $U_n^{i_2 j_2}$  describe the evolution of sizes of two subpopulations, i.e., they are rescaled versions of vertex-counting discrete processes  $V^{(1)}$  and  $V^{(2)}$ . To compute the first term in the r.h.s. of (32), we take  $0 \leq l_1 \leq l_2 \leq k$  and write

$$\begin{aligned} & \mathbb{E} \left[ V_{m+1}^{(1)} V_{m+1}^{(2)} \mid V_m^{(1)} = l_1, V_m^{(2)} = l_2, X_m = k \right] \\ &= \frac{1}{k} \sum_{i_1, \dots, i_k} (i_1 + \dots + i_{l_1})(i_1 + \dots + i_{l_2})(i_1 + \dots + i_k) p_{i_1} \dots p_{i_k} \\ &= \frac{1}{k} (l_1 B_3 + (l_1(k-1) + l_1(l_2-1) + l_1(l_2-1)) B_2 + l_1(l_2-1)(k-2)). \end{aligned}$$

Denoting

$$\begin{aligned} l_1 &= \mu n \left( \bar{U}_n^{i_1 j_1} \left( \frac{[nt]}{n} \right) \wedge \bar{U}_n^{i_2 j_2} \left( \frac{[nt]}{n} \right) \right), \\ l_2 &= \mu n \left( \bar{U}_n^{i_1 j_1} \left( \frac{[nt]}{n} \right) \vee \bar{U}_n^{i_2 j_2} \left( \frac{[nt]}{n} \right) \right), \\ k &= \mu n \bar{U}_n^{01} \left( \frac{[nt]}{n} \right), \end{aligned}$$

and making a simple calculation, we can rewrite the r.h.s. of (32) as

$$\frac{l_1}{n} \cdot \left( \frac{1}{n} + \frac{R(k, l_2)}{kn} \right),$$

where  $R(\cdot, \cdot)$  is a uniformly bounded function. We denote

$$A_n^{i_1 j_1 i_2 j_2}(t) = \frac{1}{n} \sum_{m=0}^{[nt]} \mathbf{1}_{\left\{ m \geq n(t_{i_1} \vee t_{i_2}); U_n^{i_1 j_1} \left( \frac{[nt]}{n} \right), U_n^{i_2 j_2} \left( \frac{[nt]}{n} \right) \leq U_n^{01} \left( \frac{[nt]}{n} \right) \right\}} \times \quad (33)$$

$$\begin{aligned} & \times \left( \bar{U}_n^{i_1 j_1} \left( \frac{m}{n} \right) \wedge \bar{U}_n^{i_2 j_2} \left( \frac{m}{n} \right) \right) \times \\ & \times \left( 1 + \frac{R(\bar{U}_n^{01}(\frac{m}{n}), n(\bar{U}_n^{i_1 j_1}(\frac{m}{n}) \vee \bar{U}_n^{i_2 j_2}(\frac{m}{n})))}{n \bar{U}_n^{01}(\frac{m}{n})} \right). \end{aligned} \quad (34)$$

The above calculation shows that

$$M_n^{i_1 j_1}(t) M_n^{i_1 j_1}(t) - B_n^{i_1 j_1 i_2 j_2}(t)$$

is a martingale.

We are going to use Theorem 4.1 from [EK86, Chapter 7]. Identities (31) and (33) imply two main conditions of that theorem:

$$\sup_{0 \leq t \leq 1} \left| B_n^{ij}(t) - \int_0^t b_n^{ij}(\bar{U}(s), s) ds \right| \xrightarrow{\mathbb{P}} 0,$$

and

$$\sup_{0 \leq t \leq 1} \left| A_n^{i_1 j_1 i_2 j_2}(t) - \int_0^t a_n^{i_1 j_1 i_2 j_2}(\bar{U}(s), s) ds \right| \xrightarrow{\mathbb{P}} 0.$$

The other conditions of Theorem 4.1 [EK86] are concerned with the jumps of processes  $A$  and  $B$ , they can be easily proven using the moment estimates above. The theorem allows us to conclude that the process  $\bar{U}_n$  converges weakly in the Skorokhod topology on  $D[0, 1]$  to the continuous diffusion process with drift and diffusion given by  $b$  and  $a$  respectively. In fact, Theorem 4.1 is given in [EK86] for time-independent coefficients, and here we invoke its straightforward time-dependent generalization.

Convergence of  $\bar{U}_n$  in distribution in Skorokhod topology to a continuous process implies convergence in distribution in sup norm to the same process. Moreover we can conclude that the  $U_n$ , the linear interpolated version of  $\bar{U}_n$  also converges in distribution in sup norm, which concludes the demonstration of the theorem.

## 9 Connection to superprocesses

There is an important connection of our results to the theory of superprocesses. Superprocesses are measure-valued stochastic processes describing the evolution of populations of branching and migrating particles, see e.g. [Daw93]. The limiting SPDE that we have constructed is similar to the genealogy in the Dawson–Watanabe superprocess with no motion conditioned on nonextinction, see [Eva93], [EP90], [DK99].

Our approach is more geometric than the superprocess point of view. For the superprocess corresponding to our situation, the continual mass momentarily organizes itself into a finite random number of atoms of positive mass (corresponding to discontinuities of the monotone maps in our approach). The mass of these atoms evolves in time analogously to equations (7), but our approach helps to understand what happens inside the atoms by unfolding the details of the genealogy giving rise to a contentful cocycle property whereas in superprocesses the cocycle action essentially reduces to evolving the masses of atoms. The monotone flow we construct can be initiated with zero mass and just one particle whereas in superprocesses one has to start

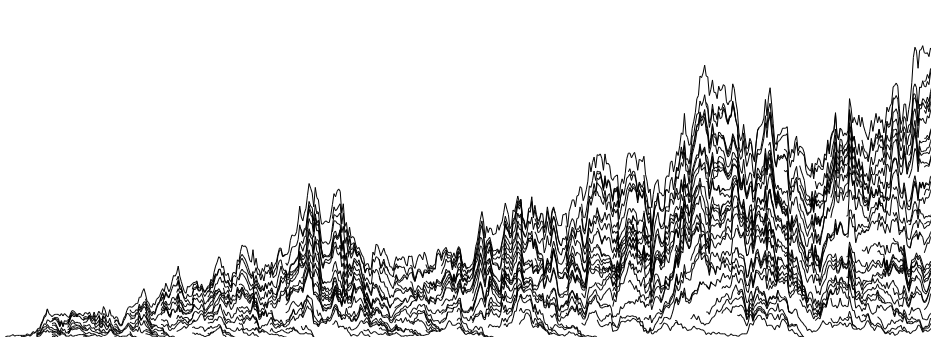


Figure 2: Stochastic foliation constructed for 600 generations of a tree.

with positive mass. The related historical superprocesses do describe the genealogy, but they require nontrivial particle motion and ignore the ordering and monotonicity issues that are central to our approach.

Notice that the continuous monotone flow we construct, describes a foliation of the random set  $\{(t, x) : 0 \leq x \leq Z(t)\}$  into diffusion trajectories. The geometry of this stochastic foliation is suprisingly complicated. Due to the presence of shocks in monotone maps (corresponding to the atoms of superprocesses), the stochastic foliation cannot be obtained, say, as a continuous image of the foliation of a rectangle in “horizontal” segments, and the flow is very far from a flow of diffeomorphisms.

We hope that our results are interesting from the point of view of graph theory since they describe what a typical large tree looks like. In fact, combining the results of this paper with those of Section 2, we conclude that a typical embedding of a large ordered rooted tree in the plane if rescaled appropriately looks like a stochastic foliation described by SPDE (13). It would be interesting to obtain rigorously a direct convergence result that would not involve the intermediate infinite discrete tree. However, currently this kind of result is not available.

Figure 2 shows a realization of a pre-limit monotone flow for a large random tree. Every tenth generation is split into about ten subpopulations, their progenies are tracked and shown on the figure.

## References

- [AP98] David Aldous and Jim Pitman. Tree-valued Markov chains derived from Galton-Watson processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 34(5):637–686, 1998.

- [Bak] Yuri Bakhtin. Thermodynamic limit for large random trees. *Accepted at Random Structures Algorithms*. Also available at <http://arxiv.org/abs/0809.2974>.
- [BH08] Yuri Bakhtin and Christine Heitsch. Large deviations for random trees. *J. Stat. Phys.*, 132(3):551–560, 2008.
- [BH09] Yuri Bakhtin and Christine Heitsch. Large deviations for random trees and the branching of RNA secondary structures. *Bull Math Biol*, 71(1):84–106, 2009.
- [Daw93] Donald A. Dawson. Measure-valued Markov processes. In *École d’Été de Probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Math.*, pages 1–260. Springer, Berlin, 1993.
- [DK99] Peter Donnelly and Thomas G. Kurtz. Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.*, 9(4):1091–1148, 1999.
- [EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes. Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.
- [EP90] Steven N. Evans and Edwin Perkins. Measure-valued Markov branching processes conditioned on nonextinction. *Israel J. Math.*, 71(3):329–337, 1990.
- [Eva93] Steven N. Evans. Two representations of a conditioned superprocess. *Proc. Roy. Soc. Edinburgh Sect. A*, 123(5):959–971, 1993.
- [Ken75] Douglas P. Kennedy. The Galton-Watson process conditioned on the total progeny. *J. Appl. Probability*, 12(4):800–806, 1975.
- [Kes86] Harry Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(4):425–487, 1986.
- [KS88] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1988.
- [LG99] Jean-François Le Gall. *Spatial branching processes, random snakes and partial differential equations*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1999.

- [Pro04] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.
- [Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.
- [Wal86] John B. Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 265–439. Springer, Berlin, 1986.